ATTACHMENT H

MONTE CARLO METHODOLOGY FOR UNCERTAINTY ANALYSIS ON THE EMBM AND FORECAST TRAJECTORY

Attachment H: Monte Carlo Methodology for Uncertainty Analysis on EMB and Trajectory Forecasts.

1.0 Introduction

Environmental systems generally have several sources of uncertainties, and these uncertainties are not merely due to a lack of proper measurements, but primarily due to the randomness inherent in real ecosystems. The implications of these uncertainties are particularly important in the assessment of several potential regulatory options, for example, with respect to the selection of a strategy for the control of pollutant levels. Incorporating these uncertainties into the modeling process, could potentially result in providing useful information that can aid in decision making.

The EMBM best estimate scenario assumed average values for all model inputs in determining the solids contribution, fate and transport of chemicals, as well as the forecast of future surface sediment concentrations under various remedial scenarios. To incorporate uncertainties in model parameters, a Monte Carlo[1] sampling approach was used to develop 10,000 iterations of each input. These 10,000 inputs were optimized in the EMBM and the optimized results were carried through the trajectory forecast calculations. A combination Microsoft Excel® Solver and the Crystal Ball® 7 (Decisioneering, Denver, CO, USA) add-on for Microsoft Excel® (a tool typically used for solving optimization problems), was used to perform this analysis. The objective of the uncertainty analysis was to provide an insight into the level of confidence in the model estimates for the best estimate scenario. This attachment presents the detailed methodology for the Monte Carlo analysis for the EMBM and Trajectory forecasts.

2.0 Methodology

The following stages were involved in the uncertainty analysis of the solids and contaminant mass balances, and contaminant forecasts presented in the CSM: (a) characterization of uncertainties in EMBM input chemical profiles, (b) estimation of the uncertainty in EMBM optimized outputs resulting from the uncertainty in chemical profiles, and (c) characterization of the uncertainties in model forecast resulting from uncertainties in the input profiles, EMBM outputs of solids contribution, decay of excess contaminant concentrations (lambda), and depth of resuspension reservoir/mixed layer (uncertainty propagation). A schematic diagram illustrating the Monte Carlo methodology is given in Figure H-1 and detailed description is presented below.

[1] Monte Carlo simulation is categorized as a sampling method in which the trails or realizations are randomly generated from probability distributions to simulate the process of sampling from an actual population.

2.1 Uncertainties in EMB chemical input profiles

Thirteen chemicals (copper, chromium, mercury, lead, trans-chlordane, 4,4'-DDE, 2,3,7,8 TCDD, total TCDD, total PCB, benzo(a)pyrene, fluoranthene, iron and TOC) were optimized in the EMBM to determine the solids balance. The uncertainties in the concentrations of these 13 chemicals for the external sources, and the resuspension source were defined by parametric and non-parametric statistics, respectively. These are described below.

2.1.1 External Sources and Receptor Profiles

The observed concentrations for the 13 chemicals for the external sources (Upper Passaic River, Newark Bay, Saddle River, Second River/SWO, Third River and CSOs) were generally normally distributed. For each external source and the receptor, a bounded normal distribution defined by the mean, standard deviation, minimum and maximum of each chemical was used to perform Monte Carlo simulation in Crystal Ball® 7. In performing these Monte Carlo simulations, it was important to maintain the relationship amongst the variables. Therefore, for each source, the correlation matrix was also specified in Crystal Ball® 7 to ensure that the 10,000 iterations of chemical profile represented the variability, inter-dependencies, and uncertainty for each external source and the receptor. Figure H-2a though H-2g presents the statistical distributions of chemical concentrations for the 13 chemicals optimized in the EMBM, for the external sources and the receptor.

2.1.2 Resuspension Source Profiles

The chemical profiles for the resuspension source were generated based on the TSI 1995 observations. The concentrations of each chemical in this data were neither normal nor log-normal distributed. None of the complex parametric distributions in Crystal Ball® 7 could adequately fit the data set. Therefore, to create the 10,000 iterations of concentrations for the resuspension source profile, a non-parametric simulation method called a bootstrap[2] was used.

The basic bootstrap approach uses Monte Carlo sampling to generate an empirical estimate of the sampling distribution of interest. In the bootstrap method, the 1995 data set was treated as the population and a Monte Carlo-style procedure was conducted on it to 10,000 iterations of the mean of the 13 chemicals optimized. This was done as follows:

1. The original sample locations of size 92 from the 1995 TSI data set were assumed to define the population of data set in surface sediments for resuspension. Note that in performing this analysis, TSI Location 246 was removed from the data set, because the PAH concentrations at this location were not representative of PAH values generally reported in the 1995 TSI data set.

[2] Bootstrap is a powerful Monte Carlo method that re-samples the original sample set with replacement to generate a distribution of sample's statistics. It is a non-parametric method.

2. The original locations were resampled with replacement to generate a bootstrap sample of size 91. This creates a bootstrap data set of the same size as the original, excluding Location 246. By resampling the locations rather than each chemical independently, the correlations amongst the chemicals were maintained. Note that this bootstrap sample set may include some sample numbers in the original sample several times, and at the same time other sample numbers may be excluded.

3. Using the chemical concentrations for the locations selected in the 91 bootstrap samples, the average concentration for each chemical was calculated.

4. Steps 2 and 3 were repeated 10,000 times to generate the empirical distribution of the resuspension source profile (Figure H-2h).

The 10,000 average concentrations generated for each chemical via bootstrap for resuspension were used along with the 10,000 iterations for the external sources and receptor to represent the uncertainty in the inputs for EMBM optimization.

2.2 Estimation of uncertainty in EMB Output

A Microsoft Excel® macro[3], which calls the SOLVER routine, was developed to perform the EMBM optimization with the aim of determining the relative solids contributions from the various sources and the mass balance for the chemicals optimized. The macro was used to solve the 10,000 optimizations using the 10,000 iterations of the sources and receptor generated by the Monte Carlo simulation. The results of the optimization run were used to understand the uncertainty in the relative source contributions and chemical mass balance for the Lower Passaic River. The 10,000 EMBM optimized results were also used as input to the trajectory forecast calculations.

2.3 Uncertainties in Trajectory Forecast

Uncertainties in forecasted chemical concentrations were defined by the results of 10,000 iterations of forecasted values. The chemicals forecasted included: 2,3,7,8-TCDD, mercury, copper, lead, 4,4'-DDE, Total PCB and gamma chlordane. Three remedial scenarios were considered including: natural recovery, remediation of Primary Erosion and Inventory zones, and remediation of sediments from RM0 to 8 (see Chapter 20 for complete description of these scenarios). Forecasting the future concentrations of chemicals under the various remedial scenarios required inputs of (i) chemical concentrations, (ii) solids contributions for the various sources determined by the EMB optimization, (iii) decay of excess contaminant concentrations

[3] A Microsoft Excel® macro is a set of instructions written in Visual Basic programming language for Application that can be triggered by a keyboard shortcut, toolbar button or an icon in a spreadsheet. Macros are used to eliminate the need to repeat the steps of common tasks over and over.

(lambda; λ), net sedimentation rate, and (iv) the depth of the sediment mixed layer. Uncertainties in these inputs were defined as follows:

1. Uncertainties in the chemical concentrations were defined by the 10,000 iterations used as inputs to the EMB optimizations (Figure H-2). For each forecast calculation, the source and receptor profiles were represented by the Monte Carlo generated values as described in Section 2.1 above.

2. Uncertainties in solids contribution from the various sources were obtained from the uncertainty in the solids contributions determined by the EMBM optimization results. This was implemented by using the 10,000 solids contribution results from the EMBM.

3. Uncertainties in decay of excess sediment contamination were defined by the regression between the natural logarithm of the excess concentrations versus time (see Figure 20-1 to 20-9 in Chapter 20). Using the slope (λ), standard error and confidence bounds from the regressions, 10,000 iterations of λ were simulated using Monte Carlo sampling from bounded normal distributions (Figure H-3). Note: $Half\ life = \ln(0.5)/\lambda$

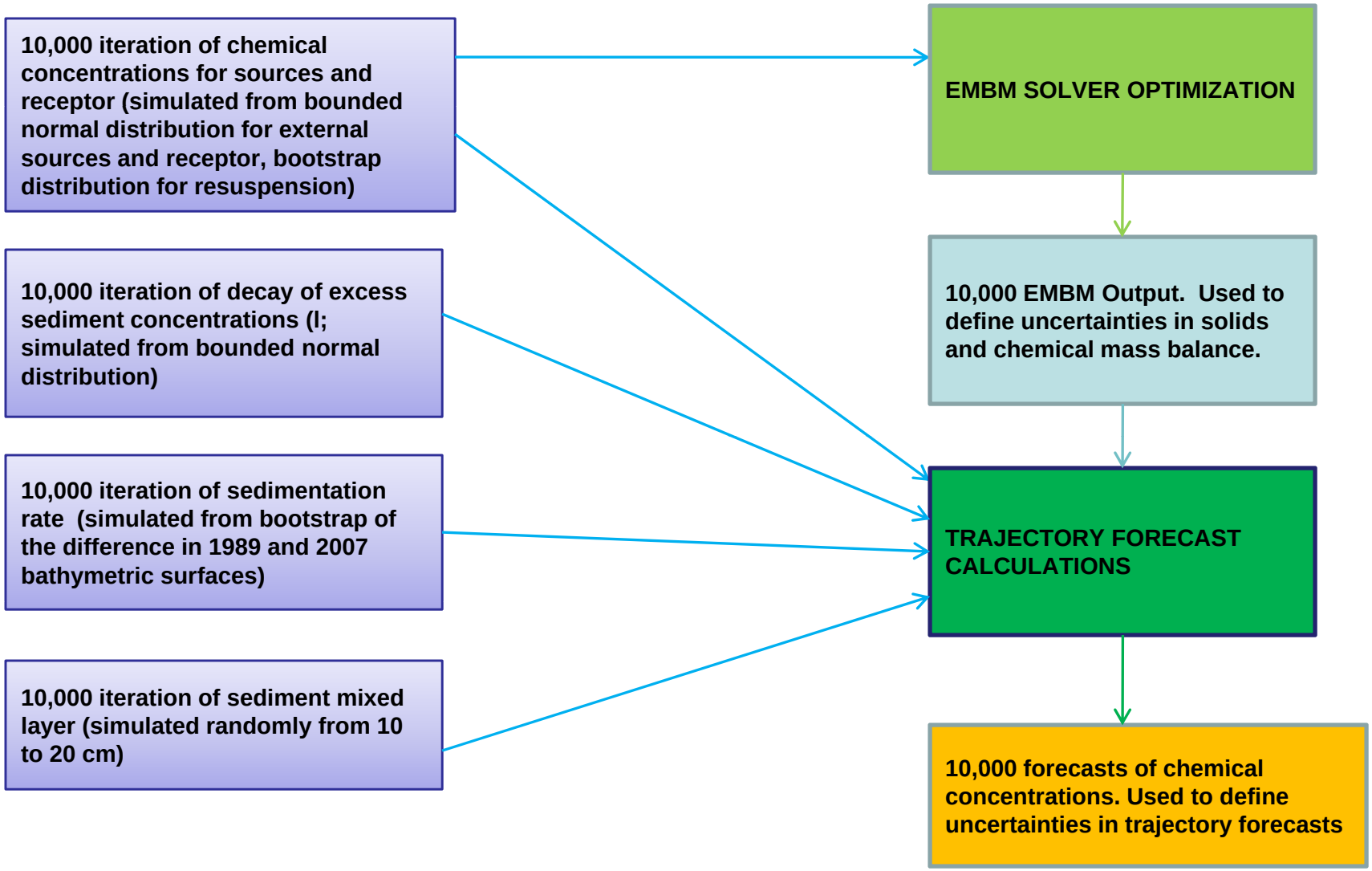
4. Uncertainties in the sedimentation rates were generated by bootstrap analysis of the differences between the 1989 and 2007 bathymetric surfaces (Figure H-4).

5. The uncertainties in depth of the sediment mixed layer were generated by 10,000 random numbers between 10 cm to 20 cm in Microsoft Excel®. Note that Microsoft Excel®'s random number generates uniform distributions of the parameter of interest (Figure H-5).

The Microsoft Excel® spreadsheets designed to perform forecast calculations using the best estimate for all inputs were modified to perform the calculations for the 10,000 iterations through a macro. For each iteration, the macro reads the input values of chemical concentrations, λ, sediment deposition rate and mixed layer depth, updates the forecast spreadsheet with these values, and then saves the results of the forecast calculation for all the remedial scenarios.

3.0 Results

Uncertainties in the EMBM solution and trajectory forecasts were defined by the confidence interval (5th and 95th percentiles) of the 10,000 optimized solutions. All the results are presented and discussed in Chapters 19 and 20.

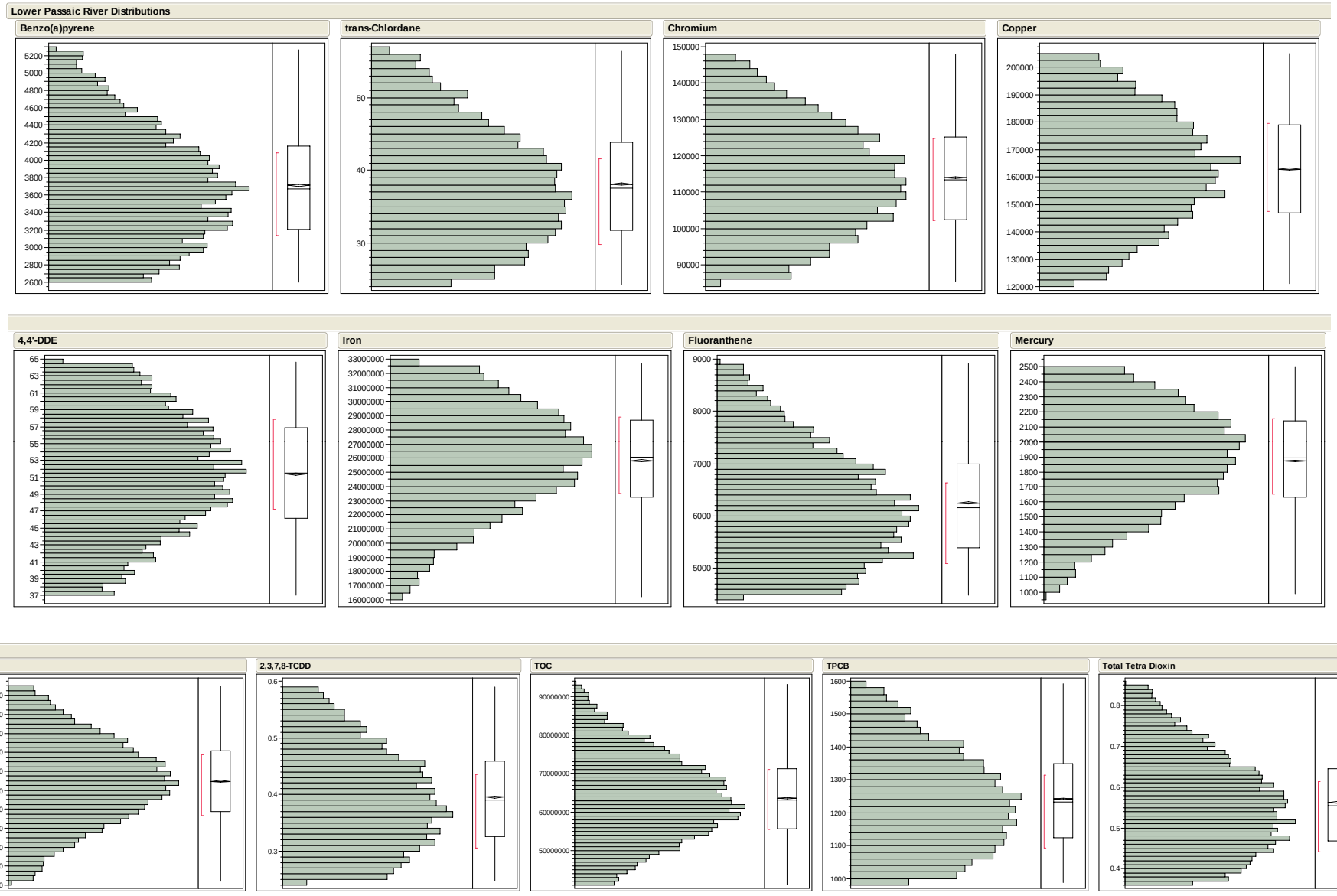


Schematic Diagram of Monte Carlo Methodology for Uncertainty Analysis on EMBM and Trajectory Forecasts

Lower Passaic River Restoration Project

Figure H-1

2009

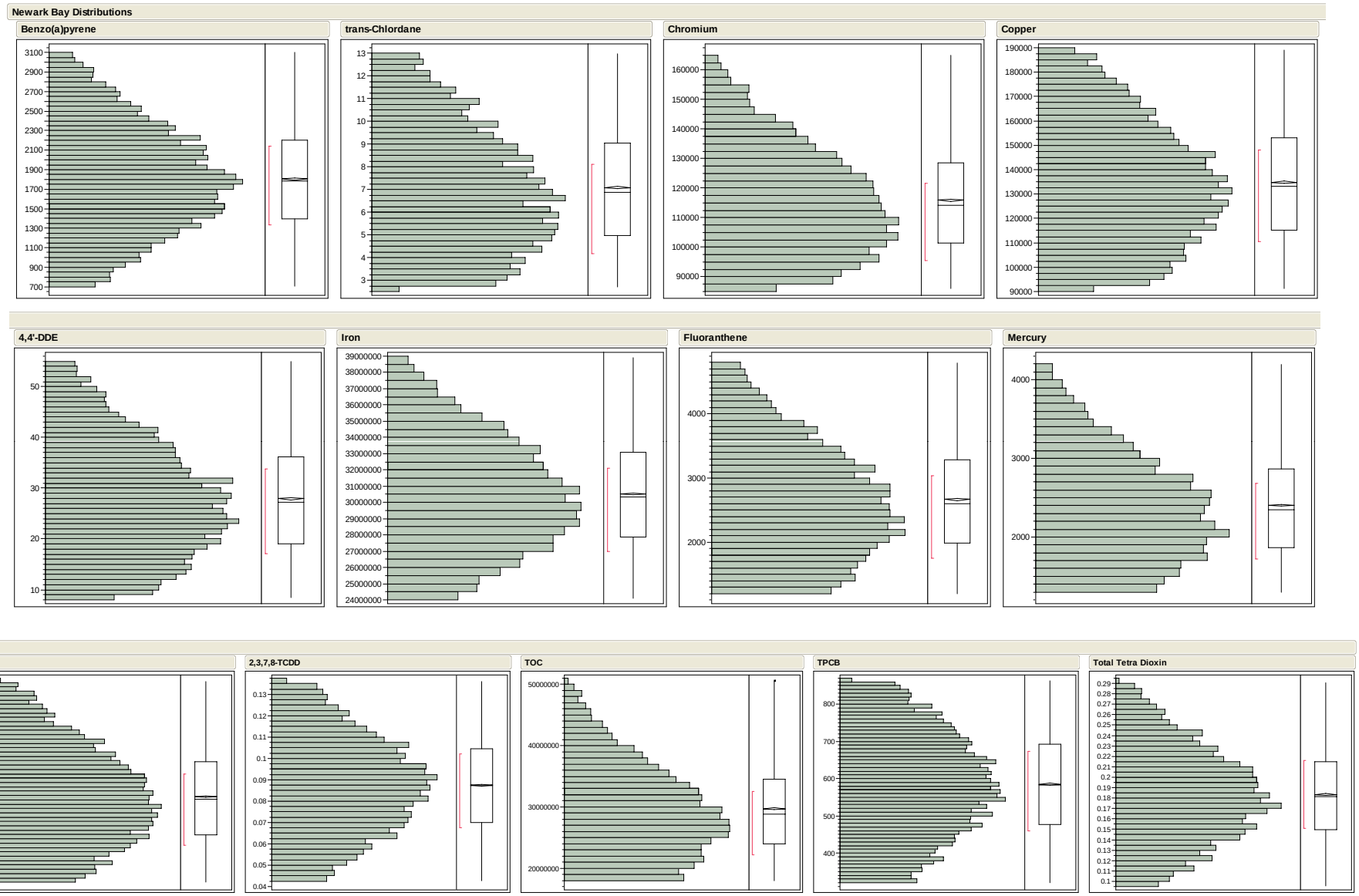


Monte Carlo Simulation of 10,000 Iterations of Chemical Profile for Lower Passaic River (Receptor) Assuming Bounded Normal Distributions

Lower Passaic River Restoration Project

Figure H-2a

2009

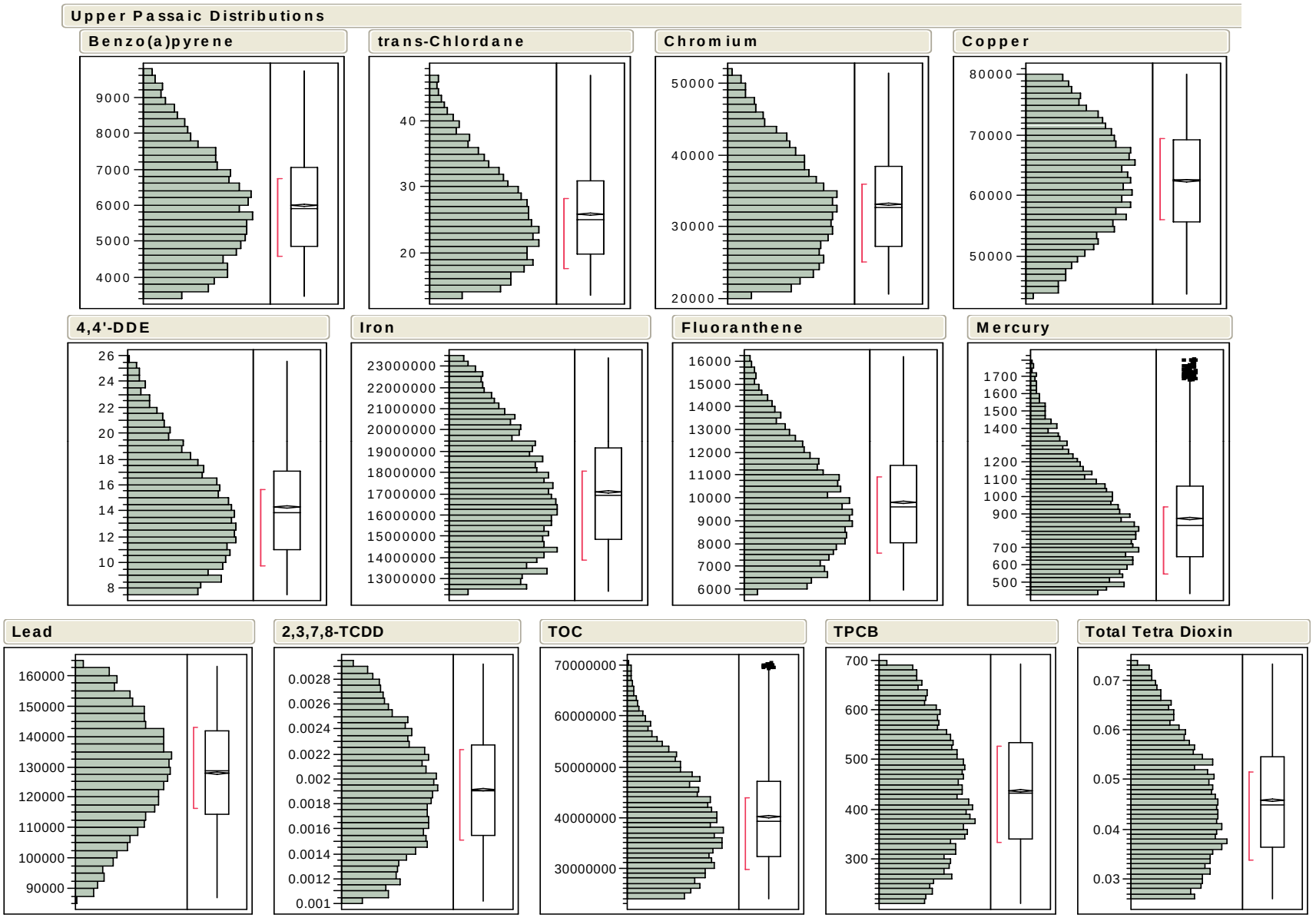


Monte Carlo Simulation of 10,000 Iterations of Chemical Profile for Newark Bay Assuming Bounded Normal Distributions

Lower Passaic River Restoration Project

Figure H-2b

2009

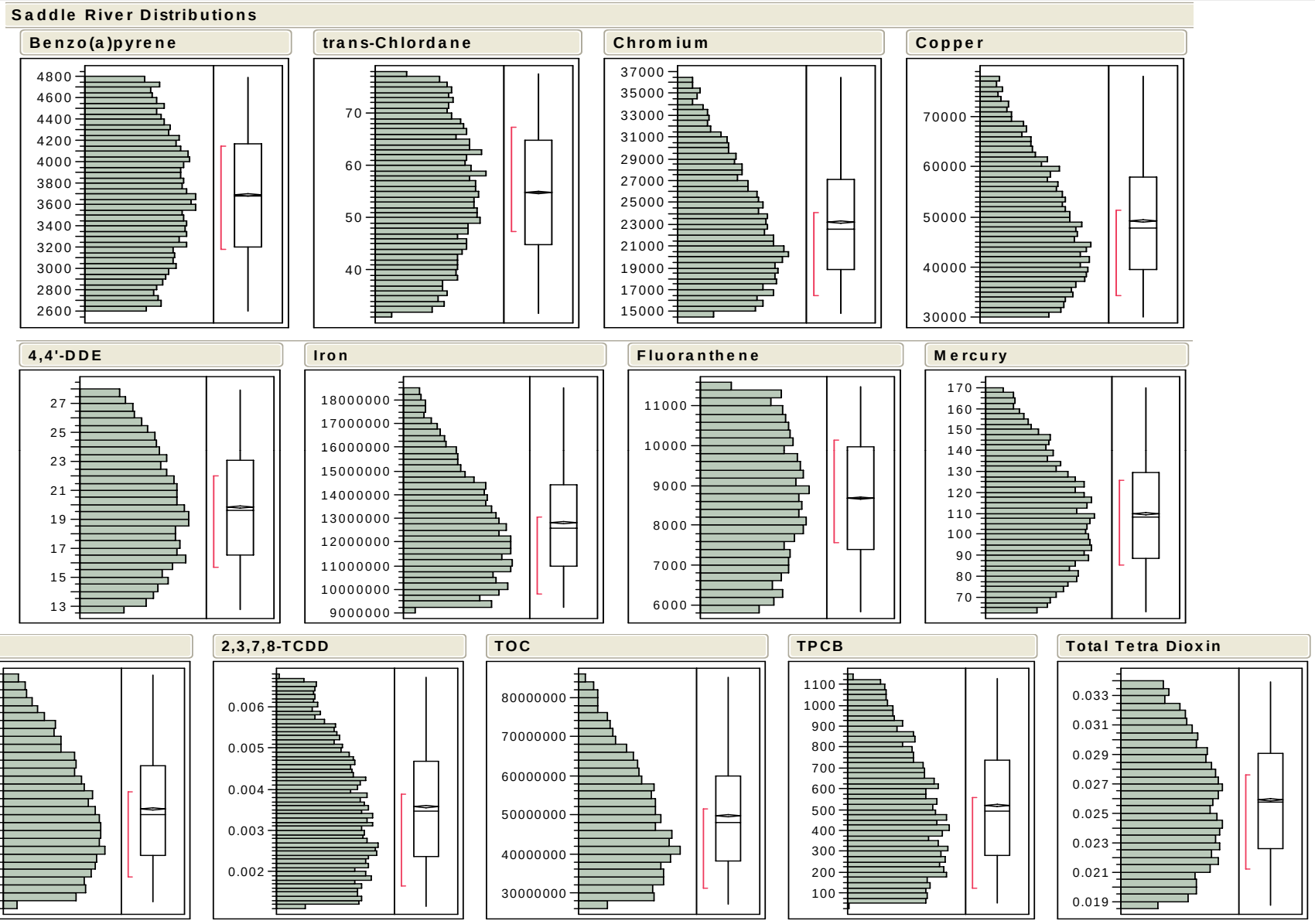


Monte Carlo Simulation of 10,000 Iterations of Chemical Profile for Upper Passaic River Assuming Bounded Normal Distributions

Lower Passaic River Restoration Project

Figure H-2c

2009

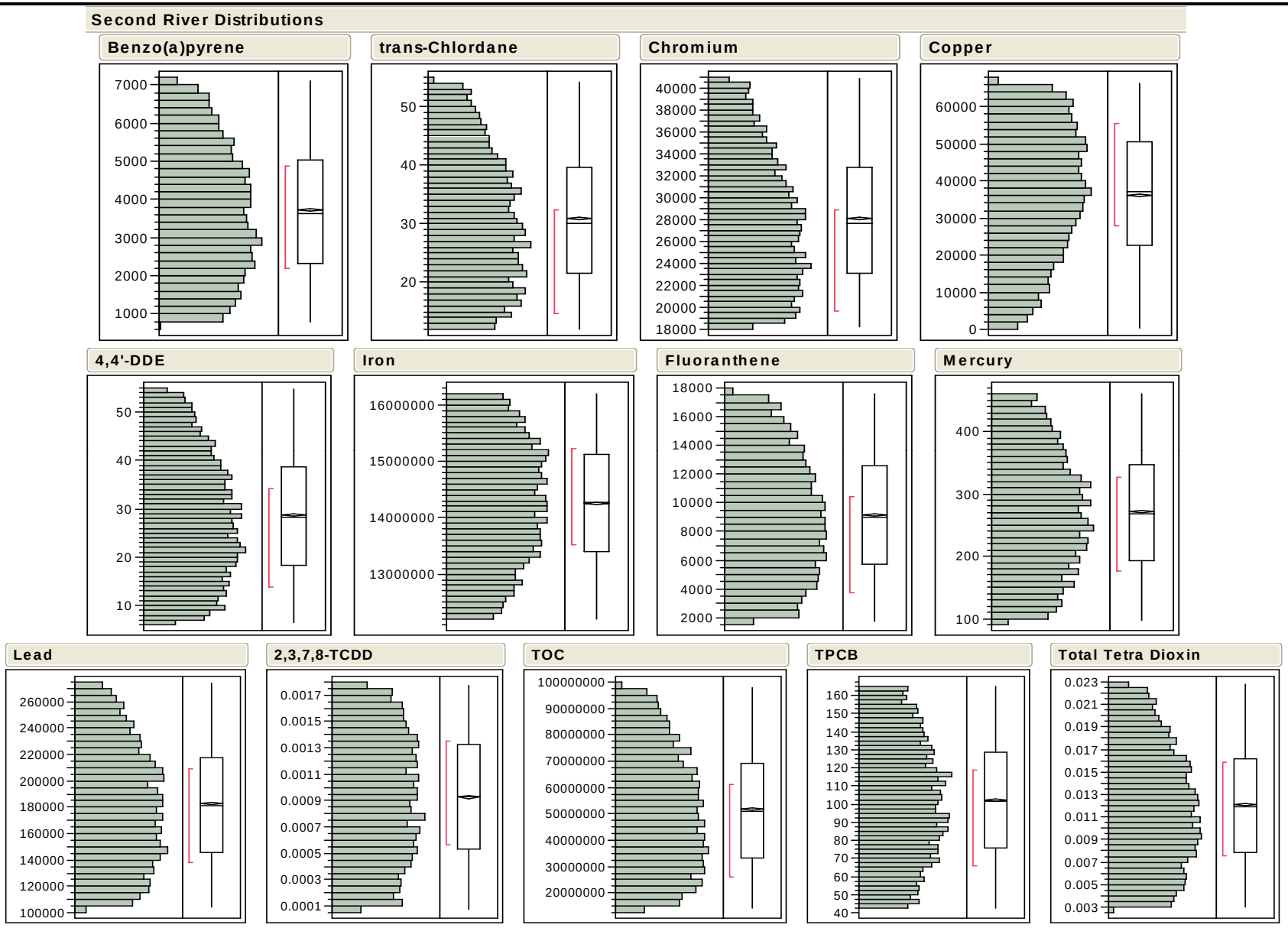


Monte Carlo Simulation of 10,000 Iterations of Chemical Profile for Saddle River Assuming Bounded Normal Distributions

Lower Passaic River Restoration Project

Figure H-2d

2009

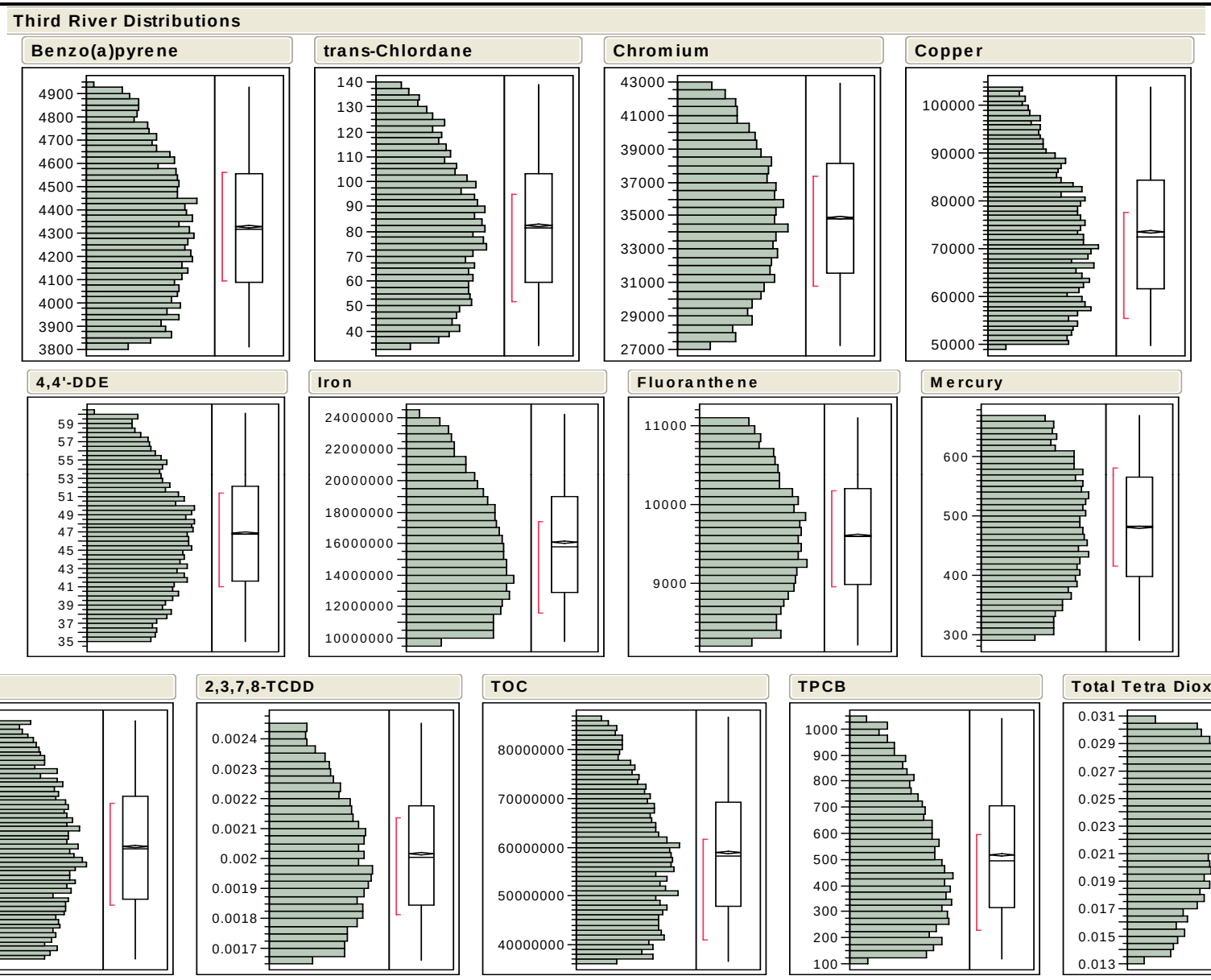


Monte Carlo Simulation of 10,000 Iterations of Chemical Profile for Second River Assuming Bounded Normal Distributions

Lower Passaic River Restoration Project

Figure H-2e

2009

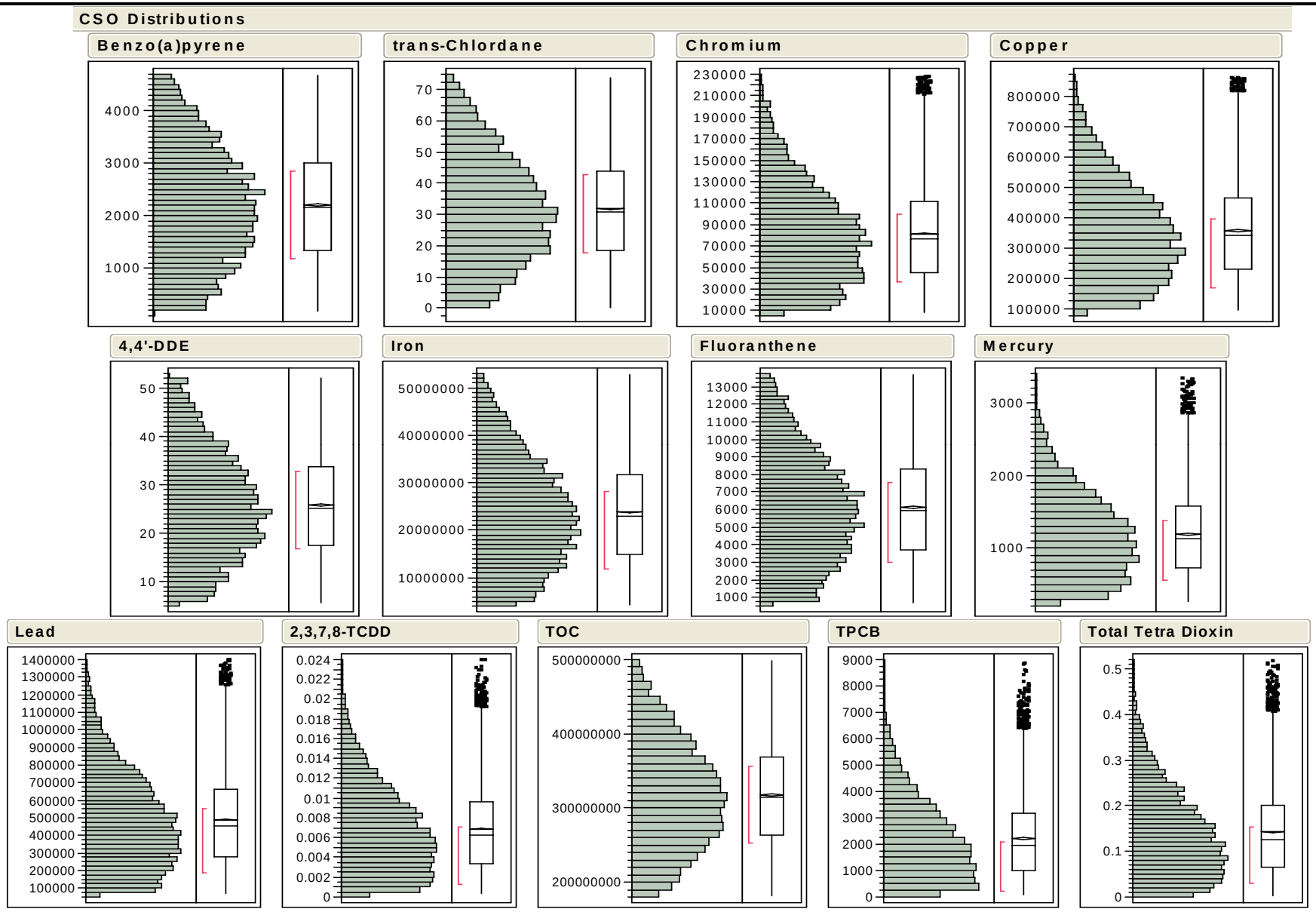


Monte Carlo Simulation of 10,000 Iterations of Chemical Profile for Third River Assuming Bounded Normal Distributions

Lower Passaic River Restoration Project

Figure H-2f

2009

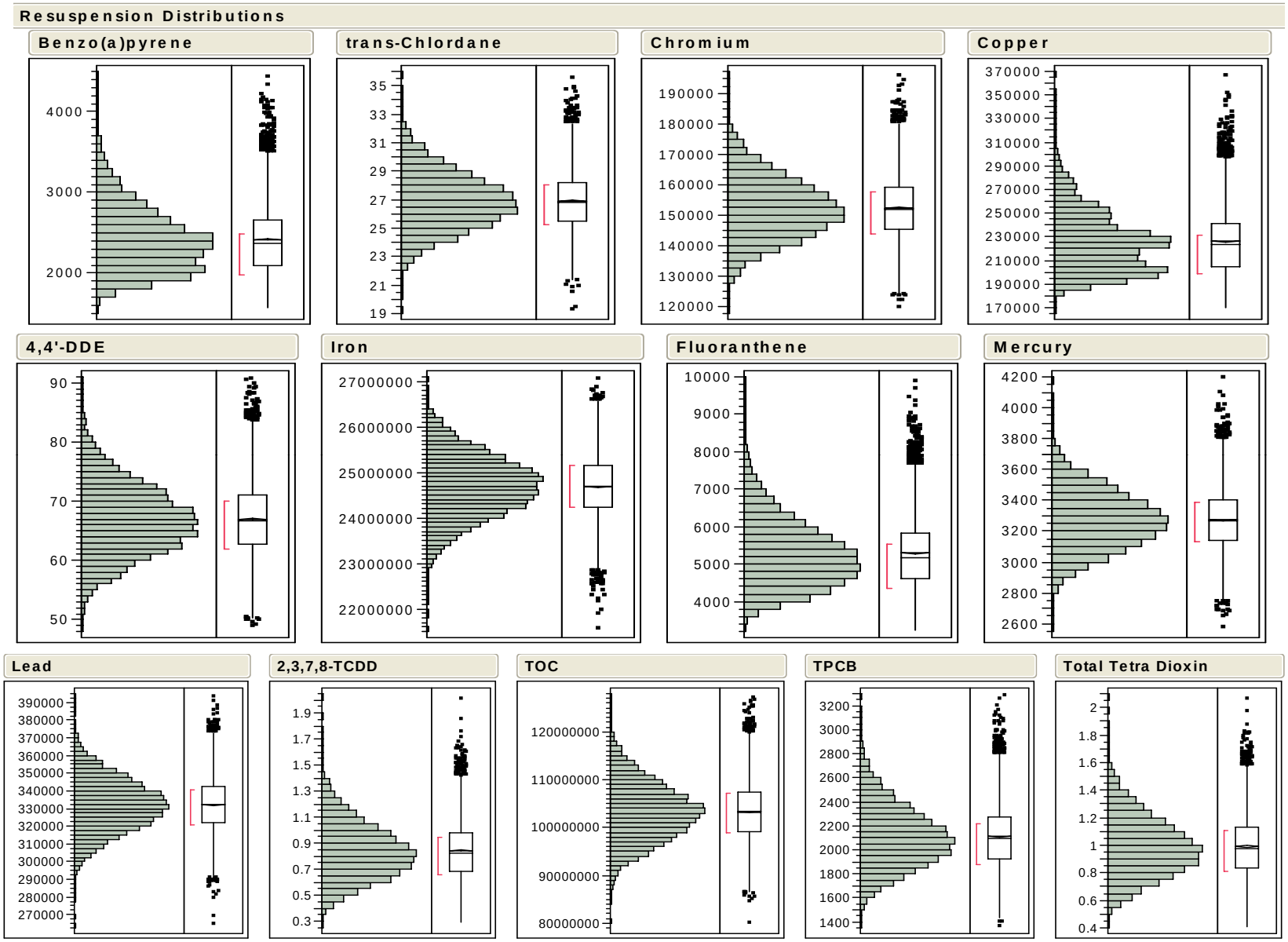


Monte Carlo Simulation of 10,000 Iterations of Chemical Profile for CSO Assuming Bounded Normal Distributions

Lower Passaic River Restoration Project

Figure H-2g

2009

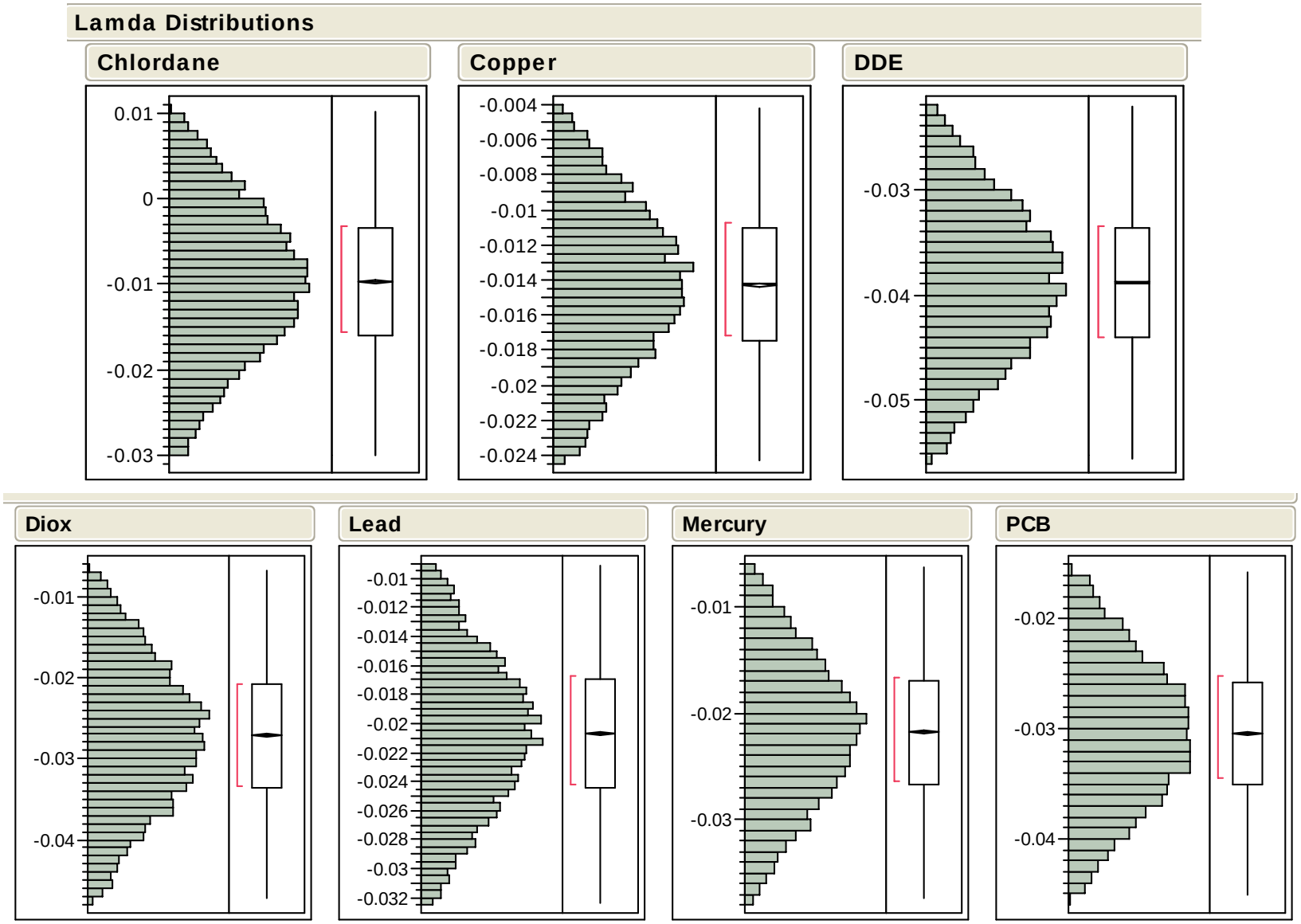


Bootstrap Simulation of 10,000 Iterations of Chemical Profile for Resuspension Source Based on 1995 TSI 0-6 inch Data (TSI Location 246 Excluded)

Lower Passaic River Restoration Project

Figure H-2h

2009

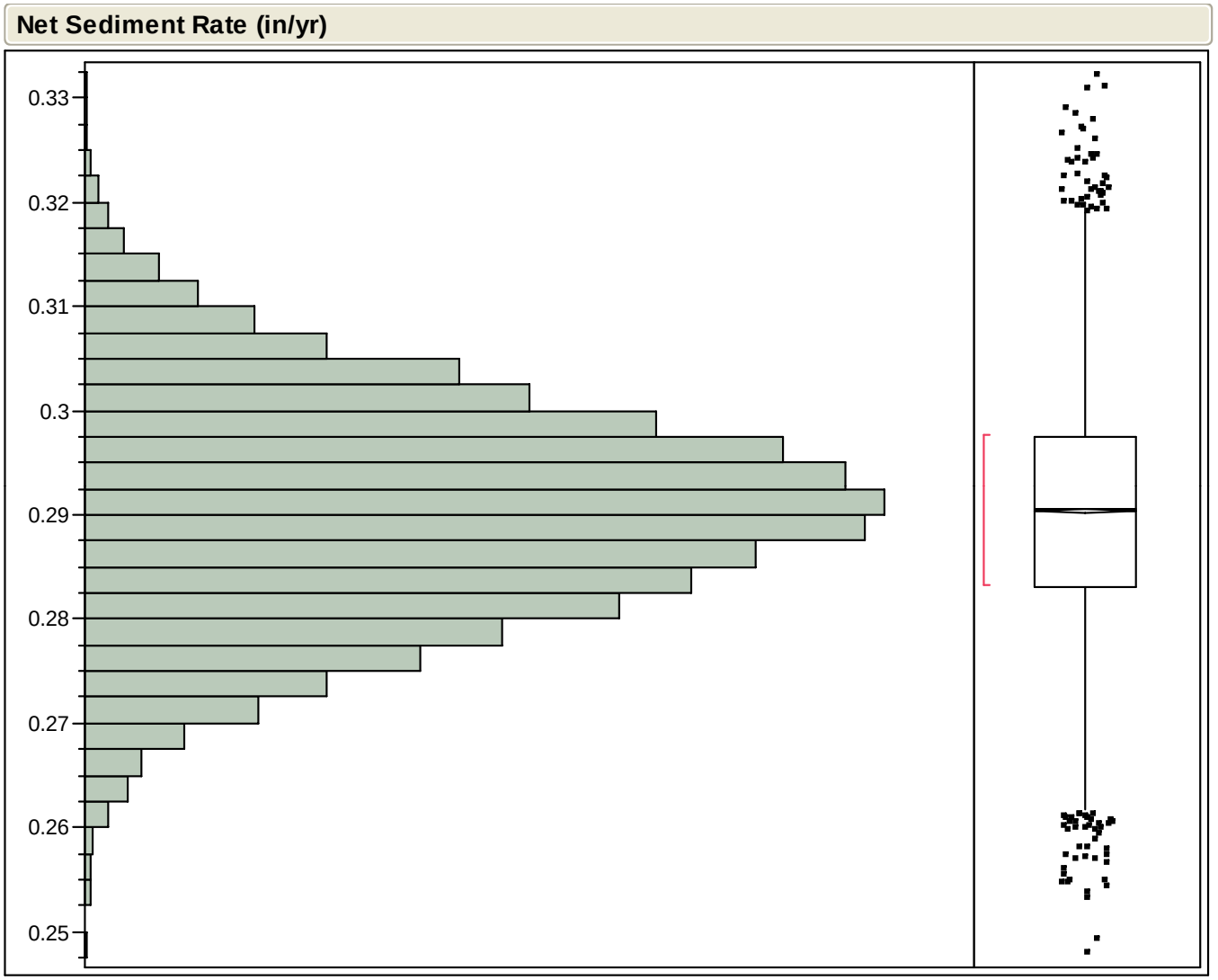


Monte Carlo Simulation of 10,000 Iterations of Decay of Excess Sediment Contamination (λ) Assuming Bounded Normal Distribution

Lower Passaic River Restoration Project

Figure H-3

2009

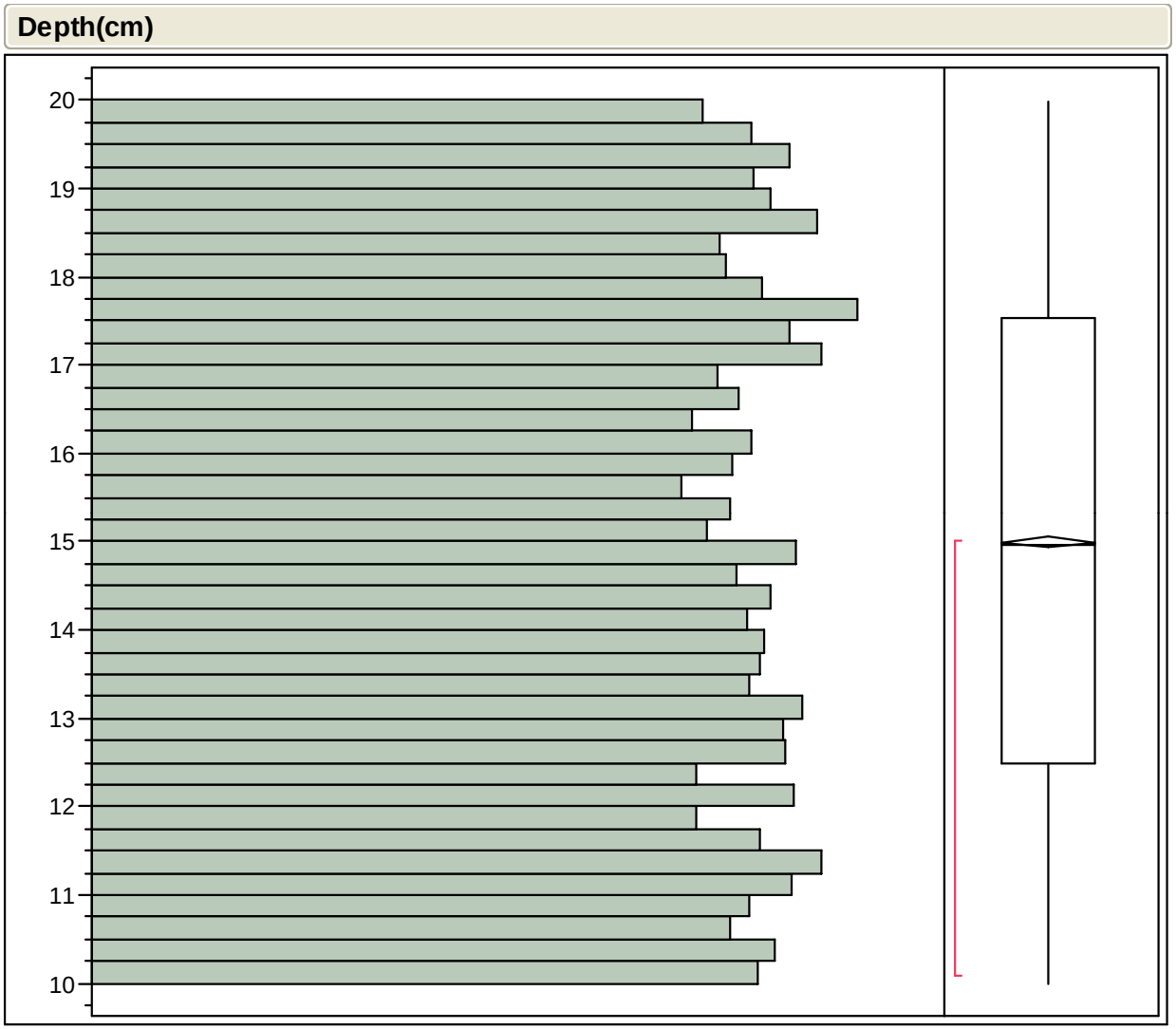


Bootstrap Simulation of 10,000 Iterations of Net Sediment Rate Based on the Differences Between the 1989 and 2007 Bathymetric Surfaces

Lower Passaic River Restoration Project

Figure H-4

2009



Simulation of 10,000 Depth of Mixed Layer Based on Random Variation Between 10 and 20 cm

Lower Passaic River Restoration Project

Figure H-5

2009

MALCOLM PIRNIE

Technical Memorandum

Date: December 22, 2008

To: Ed Garvey (NNJ)

Copy: Scott Thompson (WHI), John Kern (Kern Statistical Services)

From: S. Gbondo-Tugbawa (NNJ)

Re: Estimating the Common Half Life for Legacy Sediments in Lower Passaic River

Summary

A first-order regression model was applied to the excess chemical concentrations[1] and estimated time of deposition in the Lower Passaic River, in order to determine a common half-life for legacy contaminated sediments. The data used in the model came from the high resolution cores in the Lower Passaic River and concentrations observed for the external sources. The chemicals included in the model were: trans-chlordane, 2,3,7,8-TCDD, Total PCB, 4,4'-DDE, Mercury, Lead, and Copper. The results of the analysis indicate a common decay process[2] for these sediments at an average half-life of ~ 35 years. The 95 percent confidence interval for this common half life is from 27 to 48 years. Although only seven chemicals were included in the model, this result also applies to other particle reactive contaminants in the Lower Passaic River that have a significant resuspension source term.

Objectives

- Determine whether the chemical specific decay rates or half-lives on the excess concentrations are similar (i.e., no significant difference amongst them).
- Estimate the common decay rate for the excess concentrations in legacy sediment in the Lower Passaic River, along with the associated confidence interval.

Methods

- The chemicals included in the analysis were: trans-chlordane, 2,3,7,8-TCDD, Total PCB, 4,4'-DDE, Mercury, Lead, and Copper.
- High-resolution core data from 1980 to 2007 were used in the analysis.

[1] Excess chemical concentrations were defined as the Lower Passaic sediment concentrations less the concentrations from the external sources.

[2] The term decay is used here to quantify the net processes that result in the decline of chemical concentrations over time as observed in the high resolution cores.

- A multiple regression analysis was conducted to determine the similarities and difference amongst the half-lives of the various chemicals. This model combined the excess concentrations and time of deposition for all the chemicals. In addition, it included indicator variables for the chemical type, and allowed for interaction effects between deposition time and chemical type. The first-order regression model used was:

$$\log_e ExC_i = \beta_0 + \beta_1 T_i + \beta_2 Chl_i + \beta_3 PCB_i + \beta_4 DDE_i + \beta_5 Hg_i + \beta_6 Cu_i + \beta_7 Pb_i + \beta_8 T_i Chl_i + \beta_9 T_i PCB_i + \beta_{10} T_i DDE_i + \beta_{11} T_i Hg_i + \beta_{12} T_i Cu_i + \beta_{13} T_i Pb_i + \varepsilon_i$$

Where:

$\log_e ExC_i$ = natural logarithm of the excess chemical concentrations (i.e., high resolution core concentrations less external levels from head of tide, tributaries and CSO/SWOs)

$\beta_0 \ldots \beta_{13}$ = regression coefficients

T_i = estimated deposition time from high resolution core dating

Chl_i = indicator variable = 1 if chemical is trans-chlordane, 0 otherwise

PCB_i = indicator variable = 1 if chemical is Total PCB, 0 otherwise

DDE_i = indicator variable = 1 if chemical is 4,4'-DDE, 0 otherwise

Hg_i = indicator variable = 1 if chemical is mercury, 0 otherwise

Cu_i = indicator variable = 1 if chemical is cupper, 0 otherwise

Pb_i = indicator variable = 1 if chemical is lead, 0 otherwise

$T_i Chl_i$, $T_i PCB_i$, $T_i DDE_i$, $T_i Hg_i$, $T_i Cu_i$, $T_i Pb_i$ = interactions effects between time of deposition and chemical type

Although there are seven chemicals, only six indicators were included (indicator variable for 2,3,7,8-TCDD not included). In the statistical theory of qualitative predictor variables, a qualitative variable of c classes is always represented by c-1 indicators variables to avoid computational difficulties. In this application, the regression for 2,3,7,8-TCDD can be represented by all other indicator values being equal to zero. Note that the exclusion of the 2,3,7,8-TCDD does not affect model results. If the indicator variable of any the other chemicals modeled was excluded, the same regression results will be obtained.

- If the regression coefficients of the interaction terms are not statistically significant, then it can be concluded that the regression lines between natural

logarithm of excess concentrations versus time for the individual chemicals are parallel, and that a common decay process occurs.

Results

Table 1 presents the regression output for the first order model described above. A statistically significant model was obtained ($p < 0.001$ from Analysis of Variance results,). The most important finding from this regression analysis is that the interaction terms are not significant ($p > 0.05$). Therefore, the individual chemical regressions are parallel and there is a common decay process for the legacy contaminated sediments in the Lower Passaic River. This legacy sediment represents the resuspension source that is the dominant contribution for most chemicals. Note that the residuals of this regression satisfy the regression assumptions of normality and homogeneity of variance.

Given that a common decay process exist for the Lower Passaic River excess legacy chemical concentrations, a second regression run was conducted to estimate the common decay rate and corresponding half-life. For this regression run, the interaction terms which are not statistically significant were dropped from the regression equation. Table 2 and Figure 1 present the results for this reduced regression output. This reduced model and all the regression coefficients are statistically significant ($p < 0.0001$), and the chemical specific regressions lines are approximately parallel. The residuals of this reduced regression satisfy the regression assumptions of normality and homogeneity of variance. The regression coefficient for the time of deposition (β_1) under the reduced regression model, which represents the common decay rate is -0.02 (Table 2). This common decay rate corresponds to a half life of ~35 years. Using the standard error and t-values from Table 2 for β_1, the 95 percent confidence interval for β_1 is -0.026 to -0.014. The corresponding common half-life confidence interval is 27 to 48 years.

Table 1: Regression results with interaction terms

Multiple Regression Analysis

Dependent variable: LN_C

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	52.9204	16.7403	3.16126	0.0017
T	-0.0270905	0.00838435	-3.23107	0.0014
Chlo	-35.7128	23.5131	-1.51885	0.1299
Hg	-12.2543	21.4993	-0.569985	0.5692
DDE	23.2896	24.151	0.964335	0.3357
Pb	-12.8332	21.4993	-0.596913	0.5511
Cu	-23.8338	21.4993	-1.10858	0.2686
PCB	7.29054	23.5131	0.310063	0.7567
T_Chlo	0.0200792	0.0117754	1.70518	0.0893
T_Hg	0.00696226	0.0107712	0.646376	0.5186
T_DDE	-0.00903859	0.0120967	-0.747197	0.4556
T_Pb	0.00957072	0.0107712	0.888546	0.3750
T_Cu	0.0149399	0.0107712	1.38702	0.1666
T_PCB	0.000407156	0.0117754	0.0345769	0.9724

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1834.02	13	141.078	978.15	0.0000
Residual	39.663	275	0.144229		
Total (Corr.)	1873.68	288			

R-squared = 97.8832 percent
R-squared (adjusted for d.f.) = 97.7831 percent
Standard Error of Est. = 0.379775
Mean absolute error = 0.283226

Table 2: Regression results without interaction terms

Multiple Regression Analysis

Dependent variable: LN_C

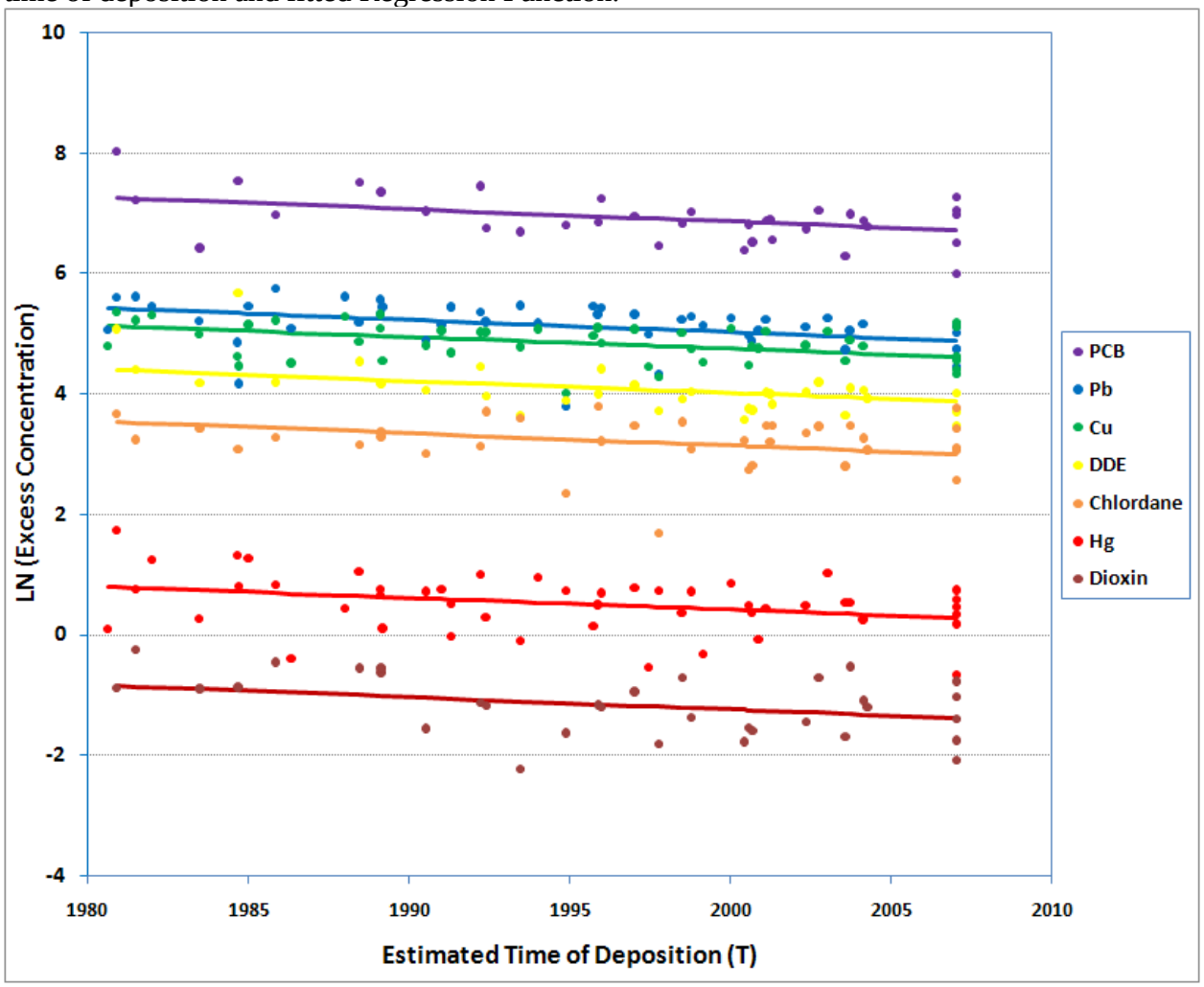
Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	38.7251	5.7386	6.74818	0.0000
T	-0.0199807	0.002874	-6.95222	0.0000
Chlo	4.38214	0.0919774	47.6436	0.0000
Hg	1.64672	0.0857078	19.2132	0.0000
DDE	5.24658	0.0932575	56.2591	0.0000
Pb	6.27178	0.0857078	73.1763	0.0000
Cu	5.98301	0.0857078	69.8071	0.0000
PCB	8.1009	0.0919774	88.0748	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	1832.76	7	261.822	1797.79	0.0000
Residual	40.9236	281	0.145636		
Total (Corr.)	1873.68	288			

R-squared = 97.8159 percent
R-squared (adjusted for d.f.) = 97.7615 percent
Standard Error of Est. = 0.381622
Mean absolute error = 0.283729

Figure 2: Illustration of natural logarithm of observed excess chemical concentration, time of deposition and fitted Regression Function.



Attachment L

Conditional Simulation on Bathymetry

Introduction

Sediment stability evaluation is a standard part of remedy selection at large contaminated sediment superfund sites. These evaluations often combine the outputs of deterministic sediment transport models, empirical mass balance arguments, geotechnical evaluations such as core dating and bathymetric change analysis. Each of these techniques has its own sets of strengths and weaknesses, but all share in a lack of flexibility in spatial scaling and uncertainty analysis that can be incorporated directly into remedy selection.

At the Lower Passaic River, bathymetric surveys were conducted in 1989, 1995, 1996, 1997, 1999, 2001, 2002, 2004 and 2007. These bathymetric surveys provide a unique opportunity for detailed temporal comparison of bathymetric changes over both long and short temporal intervals, as well as at varying spatial scales. However, direct analysis of bathymetry data presents unique statistical issues that must be carefully worked through in order to develop reliable results and interpretation.

Bathymetric survey measurements were integrated and reported on 10-foot intervals along transects and transects were spaced at approximately 100-foot intervals perpendicular to river flow. Because of this sampling design, resulting data are spatially correlated across flow and relatively variable along flow. In addition, transect locations typically varied by plus or minus 18 feet among years. These factors combined cause temporal changes of interest to be potentially confounded with cross and long flow spatial variation, as well as potential artifacts due to sample misalignment. The nature of this potential confounding is complicated by the scale of aggregation over which comparisons are made. For example comparisons of mean elevation within relatively large areas, such as hydrodynamic model grid cells are less influenced by this lack of spatial alignment than are comparisons at smaller scales.

Bathymetric change analysis typically entails interpolation of bathymetric surfaces, followed by subtraction of resulting interpolated surfaces. This approach is expected to be relatively accurate for estimation of net deposition, however, the effects of smoothing due to the interpolation can also be expected to under estimate the amount of material eroded at unsampled locations. This bias can adversely affect estimates of resuspention that are important to understanding temporal trends in contaminant concentration in surface sediments. Future risks and remedial selection are greatly dependent on understanding future contaminant concentrations in surface sediments.

Because bathymetry measurements are spatially correlated, and because sampling locations among surveys (i.e. in different years) are spatially unaligned, and because comparisons are of interest at varying spatial scales, standard statistical procedures requiring random independent samples, are unsatisfying and potentially inaccurate for some comparisons of interest. As an alternative to more conventional bathymetric change analysis methods, we used a geostatistical Monte Carlo technique

known as conditional simulation (Journel and Huijbregts 1978) to directly study the temporal change, spatial heterogeneity and spatial scale with regard to understanding sediment stability. These geostatistical techniques were selected because a probabilistic model is used to explicitly account for the effects of spatial correlation, temporal change, varying scales of interest and mismatch of sampling locations among surveys. The proposed methods are particularly powerful in that the results can be expressed as maps of the probability of sediment instability (erosion) at varying temporal and spatial scales, particularly at scales smaller than sediment transport model grid.

Geostatistical simulation procedures were developed in the mining industry in response to the need to estimate parameters and their uncertainty distributions from data that are spatially correlated. The methods explicitly provide tools to estimate parameters that vary spatially, and to propagate uncertainty over ranges of scales of interest. For example, for calibration of hydrodynamic models, the average bathymetry is of interest at the scale of the sediment transport model domain grid cells. Conversely, areas of erosion and deposition, may occur at sub-model-domain scales that if ignored could cause remedial evaluation to be biased due to the effects of smoothing the model domain. Conditional simulation provides a probability distribution at each node of a fine mesh grid (i.e. 3 foot spacing) that can then be integrated, providing error distributions for any function of the bathymetry, defined over specified areas of interest.

Conditional simulation methods have not been used widely at other superfund sites in the past, but have been recognized recently as a powerful tool facilitating analysis of uncertainty associated with metrics commonly used for remedial selection. The methods have been used widely in the mining industry, where propagating uncertainty through complicated transfer functions of geological data is important to cost analysis and reserve forecasting. As remedy selection progresses at large superfund mega sites with high remedial costs, the effort necessary to conduct a rigorous geostatistical simulation analysis of bathymetry and contaminant data will likely become more common. Recently, Kern et al (2009) used conditional simulation at the Fox River Superfund Site to refine dredge prism designs. Based on those analyses, it was found that additional sampling within planned dredging areas could be used to reduce the costs of removal of non-targeted sediments and from leaving targeted sediments behind. Recommendations of the simulation analysis were implemented in 2008, confirming that significant amounts of uncontaminated sediments were within the lateral and vertical boundaries of the 60% design dredge prisms (Barabas, et al 2009). Dredge prism refinements developed from the simulation analysis and subsequent confirming field sampling are expected to net $5 to $10 million in savings in the first year of a 10 year dredging project.

At the Passaic, we found that year over year, erosion and deposition were estimated to be nearly neutral in most areas when aggregation occurred over large spatial areas. Conversely, when considered at higher resolution (i.e. smaller scales) we found that there were many areas with relatively high probability of erosion or deposition within larger, apparently stable, transport model cells. We also used the analysis to identify areas that were erosional in at least one of the 36 possible pairs of years over the 9 year; and also to identify areas that were erosional in some years and later depositional in other years.

This attachment describes the simulation methodology used to study sediment stability in the Lower Passaic River. The results of simulated sediment dynamics from comparison of various bathymetric surfaces are presented and discussed in Chapter 11 of the CSM.

Overview of Simulation Approach

Comparison of bathymetric surveys is complicated by strong spatial correlation among soundings, within surveys, and spatial misalignment of transects, among surveys. Traditional statistical procedures such as regression and analysis of variance (Neter et. al. 1996) are designed for independent samples, or paired designs which are not possible for bathymetry surveys. These shortcomings of traditional statistical procedures were avoided with conditional simulation, because the procedures are explicitly designed for spatially correlated data and provide a mechanism to compensate for unaligned sample locations.

The conditional simulation approach is based on the same ideas as a Monte Carlo or parametric bootstrap analysis where inference is based on the ensemble properties of a large number of random draws from a probability distribution thought to have the statistical characteristics of the properties of interest. The primary difference is that the underlying statistical distributions are assumed to be spatially correlated, and the random draws are from a population of maps rather than populations of individual values. The population of maps are consistent with the statistical distribution of sample data, have the same spatial correlation relationships, and also interpolate the sample data. This method has also been referred to as *"stochastic interpolation"*. Inferences are developed by repeatedly drawing equally likely maps, calculating transfer functions of the mapped values, such as erosion and repeating the process. The result is a histogram of the output functions of interest at selected scales of aggregation. In this way, it is possible to construct confidence intervals of erosion at essentially any selected scale while properly incorporating uncertainty due to sampling error, misalignment of transects and spatial correlation.

The analysis required several data handling procedures, including; 1) Grid Straightening, 2) Detrending, 3) Semivariogram analysis, and 4) calculating sediment erosion and deposition at scales of interest.

Grid Straightening

Geostatistical procedures require estimation of the spatial relationships among sampling locations in order to determine how best to simulate or interpolate to unsampled locations. In typical geological and geographical settings, these relationships are measured along geographic coordinate axes and summarized by the semivariogram—a function that quantifies variance as a function of distance. For positively spatially correlated processes, variance is an increasing function of distance. In other words, dissimilarity is expected to increase with distance between sample locations. However, when studies involve data collected within riverine or estuarine systems, spatial relationships are determined largely by flow directions. In general, it is expected that sample locations along similar flow paths would be

more similar in value than those that are located across flow directions. For bathymetry this can be seen by the obvious orientation of depth contours that parallel the direction of flow.

Because the Passaic River is sinuous, variogram analysis based on geographic coordinates would not properly capture the spatial relationships that are expected to be driven by flow directions. To correct this situation, the geographic coordinates were transformed to long- and cross-flow coordinates using the Schwarz–Christoffel Toolbox (Driscoll, 1996) implemented in MATLAB© (The Math Works, 1998).

The coordinate transformation (Figure 1) was done as follows: 1) the physical boundary was defined by the two banks and lines marking the northern (RM 7.5) and southern (RM 0.5) ends; 2) curvilinear grid lines parallel and perpendicular to the flow were constructed using the Schwarz–Christoffel conformal mapping; 3) the Schwarz–Christoffel Toolbox was used to transform all data locations from rectangular (Euclidean) coordinate to the constructed curvilinear system. One unit of distance corresponds in the new curvilinear scale to an average of ~150 ft in the Euclidean scale. All geostatistical analyses were performed in the transformed coordinate system.

Exploratory Data Analysis and Transformations

The basic assumptions of the geostatistical model that we applied are that the bathymetric elevations can be mathematically transformed to a population that;

1. has constant mean and variance across the area of interest,
2. spatial correlation that is a function of distance and direction, and
3. a normal distribution.

One and two are the assumptions of second order stationarity (Cressie 1991) and additional of assumption three leads to the multigaussian model described by Deutsch and Journel (1992) and also by Goovaerts (1997) with applications in natural resources characterization. This assumption is similar to the assumption of constant variance in a traditional regression analysis. The purpose of the transformation and detrending of the bathymetry data that we describe below is analogous to the variance stabilizing transformations commonly applied prior to a regression analysis.

Histograms

Figures 2a through 2h show the histograms and summary statistics of the bathymetry data from each survey. It can be seen that the distributions are somewhat left skewed, and tend to have an over abundance of measurements between 0 and -5 feet below sea level. This deviation from normality is expected to require some level of transformation prior to application of the multi-Gaussian model for simulations. We used the normal score transformation to correct for non-normality.

Detrending

For each survey, it was found that the cross-flow semivariograms have much shorter ranges of influence and also much higher maximum values. This is an example of zonal anisotropy (Journel and Huijbregts 1978; p 182) caused by directional differences in variation. It is not surprising that elevations would vary more rapidly across flow than along flow due to the quasi-U shaped profile of the river cross sections. In

geostatistical terms, this is indicative of a non-stationarity of the mean that could adversely affect inferences if not modeled carefully. In order to minimize the potential effects of strong zonal anisotropy on simulation results, the data were detrended.

Bathymetry elevations in the Passaic River exhibit a strong U-shaped channel profile with a thalweg that meanders within the channel tending to be closer to outside bends in areas of higher curvature. Elevations also tend to decline gradually with distance downstream. These large scale trends would violate the constant mean assumption described above for the geostatistical model, so these large scale patterns were estimated and subtracted from the original elevations so that the resulting residuals could be treated as a second order stationary process. Detrending is a common approach when analyzing spatial data (Cressie, 1990 p. 46) and is part of the process of partitioning data into large and small scale fluctuations that are of interest. In this application, the primary interest is in understanding relatively small (3-6 inch) changes in elevation among surveys, so larger scale spatial variation associated with channel cross sections are would tend to mask these smaller fluctuations of primary interest.

Large scale spatial patterns in the bathymetry are well understood—including an approximately U-shaped river cross section and a general downward sloping trend in the long-flow direction. It is also understood that the position of the minimum elevation in each cross section (the thalweg) varies making it difficult to model using a simple polynomial function of the geographic coordinates. Polynomial regressions were tested with little success resulting in very low R-squared values (less than 20 percent; Figure 2i-2p) and the residuals also show strong zonal anisotropy (Figure 2j). As a substitute for more conventional trend surface approaches, the large scale trend was estimated by calculating the average elevation within relatively large rectangular grid cells, and then smoothing the estimated cell means with a moving window average (Isaaks and Srivastava, 1989). An example of the resulting trend surface for the 1995 bathymetry can be seen in Figure 3. The objective of detrending was to partition the bathymetry into a large scale trend component and a second order stationary residual process. The size of the rectangular cells was selected by iterating on the semivariograms, starting with relatively large cells and decreasing them until the zonal anisotropy in the semivariograms was eliminated—consistent with the assumptions of constant mean and variance.

Semivariogram Analysis

Bathymetry elevations were detrended as described above and semivariograms were recalculated for each survey. The P-field simulation algorithm used for these studies requires semivariograms for both uniform and normal score transformed data (Srivastava, 1992). Therefore the detrended residuals were transformed to uniform and normal scores and semivariograms were calculated for each. The resulting empirical semivariograms and fitted positive definite models can be seen in Figures 4a-4p. It can be seen that the long- and cross-flow sills are similar for these directional variograms, indicating that the zonal anisotropy has been greatly reduced or eliminated by detrending. Semivariograms were constructed using the *gamv* procedure in GSLIB (Deutsch and Journel, 1998).

Kriging Analysis

Another intermediate step in the simulation process is development of a Kriged estimate of the mean and kriging variance for each unsampled location using the normal scores transformed residuals and corresponding semivariograms. This analysis was conducted using the kt3d procedure in GSLIB (Deutsch and Journel, 1998). The simple kriging option was selected. The estimated mean and kriging variances were used to estimate the conditional cumulative distribution at each location under the assumptions of the multi-Gaussian model. In effect the cumulative conditional distribution is assumed to be Normal with means and variances given by the Kriged estimates.

Simulation Algorithm

Conditional simulation has been understood to be a potential tool for investigating uncertainty associated with remedial alternatives selection and design at large complex contaminated sediment sites, but the better known algorithms available to practitioners, such as sequential Gaussian simulation (Ripley, 1987; Deutch and Journel, 1998), tend to be extremely computer and time intensive, often limiting their application. In the mid 1980s to 1990s, methods were developed at the University of Wyoming (Borgman et al. 1984) that reduced the necessary computational time of spatial simulation through application of the fast Fourier transform (FFT). Kern and Borgman (1997) described the algorithm in detail, demonstrated the accuracy of the method and compared it with a sequential algorithm for reproduction of second order moments and computational speed.

The frequency domain methods are so termed, because spatially correlated data are transformed to a series of two dimensional sequences of sine and cosine functions (i.e. their frequency components) which are statistically independent. Because the Fourier coefficients are statistically independent, the space domain simulation problem of generating many correlated variables with constant variance is reduced to simulation of a vector of independent random Fourier coefficients with frequency dependent variances. This vector of independent random variables is then "reorgainized" using the inverse FFT to produce a space domain simulation with the specified covariance relations. Effectively, the sequential procedure is replaced by one application of the inverse FFT to the full vector.

The method is faster in very practical terms. For example, when simulating values at N spatial locations, the computational time for the more widely known sequential simulation procedures (SGSIM; Deutch and Journel, 1998) is proportional to N^2, while the computational time for the FFT method is proportional Nxln(N), so for large simulation problems, such as for the bathymetry analysis where N > 1,000,000 spatial locations, the FFT method reduces computation time dramatically. A single simulated realization of 1,000,000 spatial locations requires approximately 2.6 seconds with the FFT method. A similar calculation using SGSIM or other sequential algorithms would require approximately 45 minutes per realization. In practical terms, including post processing of simulation results, the significant reduction in computational time of the FFT approach dramatically increases the feasibility for simulation studies that can be handled within reasonable time frames. After all, conditional simulation is a Monte Carlo technique requiring many realizations to provide accurate inference. Depending on the size of the spatial domain, the FFT method can be used to generate many hundreds to thousands of realizations in

minutes to hours, while the competing algorithms would require days. In this study 500 realizations from each survey were constructed and post processed.

One might ask how many realizations are adequate to estimate the things like the probability of erosion we have used in this application. A rough way to consider this question is to consider that probabilities associated with each cell in the simulation grid could be treated as proportions from a binomial distribution. If one defines *"p"* to be the estimated proportion of realizations in which erosion was observed, the standard error of the estimated *p* is given by $se(p) = \sqrt{p(1-p)/K}$, where *K* is the number of realizations. So for estimated probabilities near 0.7 the standard error for *K*=500 realizations is 0.02. This indicates that if the simulations were repeated using the same histogram and semivariogram parameters, one would expect probabilities ranging from 0.68 to 0.72.

P-Field Algorithm

Direct conditional simulation with the FFT method as described by Borgman (1984) is feasible for small numbers of conditioning data (i.e. sampled locations) relative to the number of simulation nodes. A typical desktop or laptop computer with 3G of memory can reasonably handle up to approximately 1000 sample points and one million simulation nodes. Bathymetry data are unique among other types of environmental data in that there are typically hundreds of thousands of sample locations for conditioning, so the direct FFT approach would not be feasible without dramatically larger capacity computer resources. To work around this problem an alternative simulation algorithm known as P-Field (Srivastava, 1992) was selected because the algorithm can be combined with the FFT, resulting in an accurate simulation algorithm with speed and efficiency similar to the full FFT approach.

The P-Field algorithm includes estimation of the CCDF at each of the unsampled locations conditional on the observed sample data through kriging. Once these CCDFs are constructed, each realization is sampled by selecting a uniform random value at each location and inverting the CCDF evaluated at the selected uniform value to produce a simulated value in the original scale. The idea behind the P-Field algorithm is that the uniform variables are spatially correlated to insure that the values drawn from the CCDFs reproduce the spatial correlation of the original untransformed data. Simply selecting uniform random values for each location without regard to spatial correlation would result in simulated realizations that reproduced the sample histogram, but with inflated nugget effect. So the FFT algorithm is used to produce unconditional realizations of the uniform random variables that are properly spatially correlated. This step in the process is the most computationally intensive for a sequential algorithm, requiring minutes per each of the 500 realizations, whereas the FFT method requires just 3 to 6 seconds per realization. The method is extremely fast and each realization reproduces the original sample histogram and semivariogram.

Surface Comparisons

To compare bathymetric elevations among years, an equally likely surface was simulated for each year of interest, and the elevations were subtracted on each pixel in the simulation grid. These differences

were compared with difference cutoffs of interest, such as 3-inches, 6-inches or 12-inches and the number of times a simulated difference exceeded each cutoff value was recorded for each pixel. These frequencies were then divided by the total number of simulations (K=500) representing the probability of erosion at these selected depths of interest. Differences were also integrated over areas of interest, such as transport model grid cells, providing probability distributions for the amount of erosion averaged over these cells. The results of these and other comparisons are provided in Chapter 11 of the conceptual site model.

Figure 5 provides three realizations of the simulated elevation surface for 1995. It can be seen that the sample data constrain the range of variation, yet there are distinct differences among simulated surfaces that represent the uncertainty remaining in bathymetry interpolation.

Figure 6 illustrates the histograms of simulated and actual bathymetric data. It can be seen that the simulation algorithm reproduces the input sample histogram as desired. It should be noted that this is a comparison of the histogram of all simulated locations to the data histogram, as opposed to the subset of simulated locations at which comparisons were constructed. This latter comparison cannot be developed without biasing artifacts of sampling, because the distribution of samples is not independent of the subset of cells used for comparison, because the comparison subset is restricted to deeper areas than that represented by the sample data. Because of this biasing one would not expect the sample histogram to match the simulated histogram within a sub-area of the river.

Figure 7 shows the comparison between the theoretical long- and cross-flow semivariogram models and actual semivariograms calculated from 20 simulated surfaces. It can be seen that the semivariogram is reasonably well reproduced for both directions. If anything, these 20 realizations may slightly under-state the long-flow variogram which would tend to result in understatement of the magnitude of long range fluctuations and potentially creating a small degree of smoothing in the long-flow direction.

Discussion

This attachment is intended to document the simulation procedure that was used to support analyses reported in Chapter 11 of the CSM. The methods are relatively new to the environmental sciences, but are well documented in theoretical and practical developments in the geological engineering and geostatisitics literature. The advanatages of the conditional simulation approach are worth-while when data are correlated and inferences are needed at scales other than the original data. Essentially, any complicated non-linear transfer functions of the surface can be accommodated. In this instance, separate estimates of volume of material eroded (deposited) were of interest and interpolated maps did not provide adequate understanding of the magnitude and frequency of erosion at sub-model grid scales. These methods provided an improved understanding of these processes.

References

Borgman, L.E., M. Taheri, and R. Hagan. 1984. Three-dimensional frequency-domain simulations of geological variables. In G. Verly, M. David, A. G. Journel, and A. Marechal, Editors, *Geostatistics for Natural Resources Characterization,* volume 2, pages 517-541.

Barabas, N., J.R. Wolfe and J.W. Kern. 2009. The influence of uncertainty in models of dredged sediment volumes vs. sampling density. . Fifth International Conference on Remediation of Contaminated Sediments. February 2-5, 2009 Jacksonville, Florida.

Deutch, C. V. and A.G. Journel. 1998. *GSLIB: Geostatistical Software Library and Users Guide.* Oxford University Press, New York, 370 p.

Driscoll, T. A., A MATLAB toolbox for Schwarz–Christoffel mapping, ACM *Trans. Math. Soft*., 22 (1996), pp. 168–186.

Goovaerts, P.M. 1997. *Geostatistics for Natural Resources Evaluation.* Applied Geostatistics Series, Oxford University Press, New York. 483p.

Journel, A.G. and Ch. J. Huijbregts. 1978. *Mining Geostatistics.* Academic Press, New York. 600 p.

Kern, J.W. and L.E. Borgman. 1997. Frequency domain conditional simulation using the fast Fourier transform and recently developed consistency conditions. United States Department of Interior, Office of Surface Mining Reclamation and Enforcement, Technology TYransfer. http://www.techtransfer.osmre.gov/NTTMainSite/Library/pub/fdcs.pdf. (Last visited 4/19/2009).

Kern, J.W. J.R. Wolfe and N. Barabas. 2009. Sampling Density for Refinement of Dredge Prism Design in the Lower Fox River, Wisconsin OU 2-5. Fifth International Conference on Remediation of Contaminated Sediments. February 2-5, 2009 Jacksonville, Florida.

Neter, J., Kutner, M.H., Nachtsheim, C.J. and W. Wasserman. 1996. *Applied Linear Statistical Models.*Fourth Edition. IRWIN Chicago, Ill.

Ripley, B.D. 1987. *Stochastic Simulation.* John Wiley and Sons, New York, NY.

Srivastava, R.M. 1992. Reserve characterization with probability field simulation. In SPE *Annual conference and Exhibition, Washington, D.C.,* Number 24753, pages 927-938, Washington D.C., Society of Petroleum Engineers.

Srivastava, R.M. and R. Froidevaux. 2005. Probability field simulation: A retrospective. In: Leuangthong, O. and C.V. Deutsch (eds). *Geostatistics Banff 2004.* 55-64. Springer. Netherlands 2005.

The Math Works. 1998. MATLAB, The Language of Technical Computing, Version 5.2. www.themathworks.com. Last visited June 19, 2009.

Figure 1: Long flow and cross flow coordinate axes resulting from the Schwartz-Christoffel conformal mapping in a selected part of the lower Passaic River.

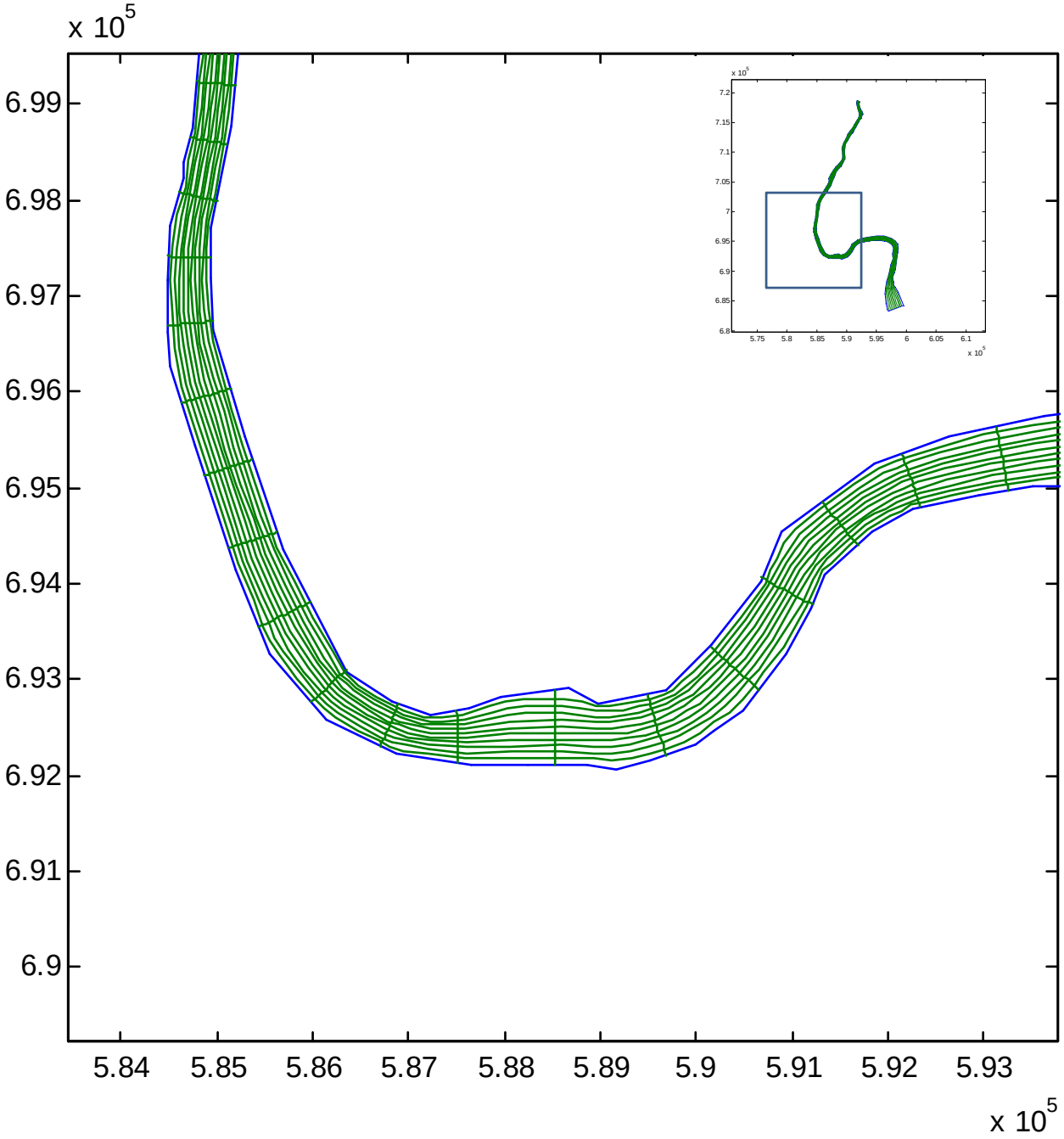


Figure 2a. Histogram and Summary Statistics for 1989 Bathymetry Survey

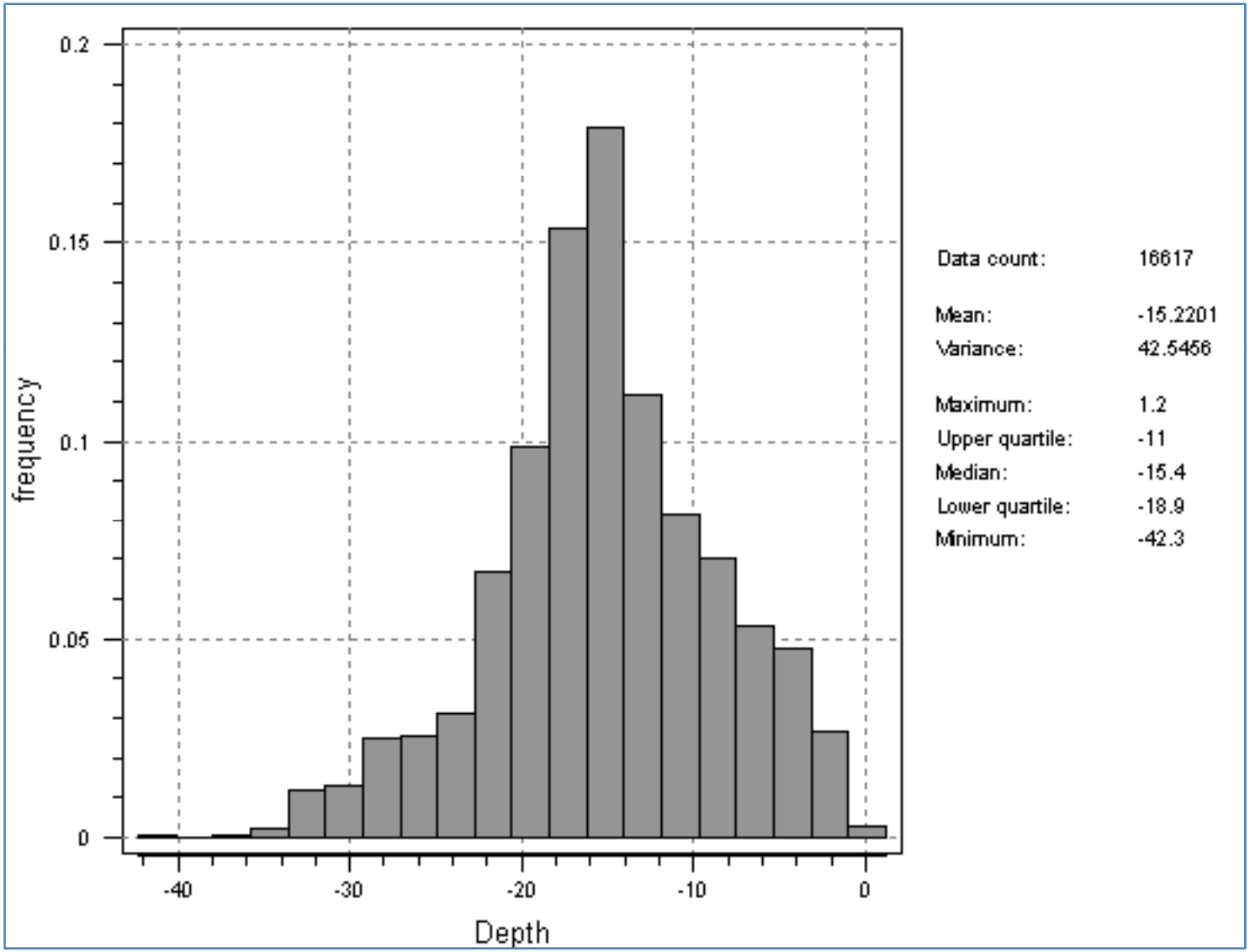


Figure 2b. Histogram and Summary Statistics for 1995 Bathymetry Survey

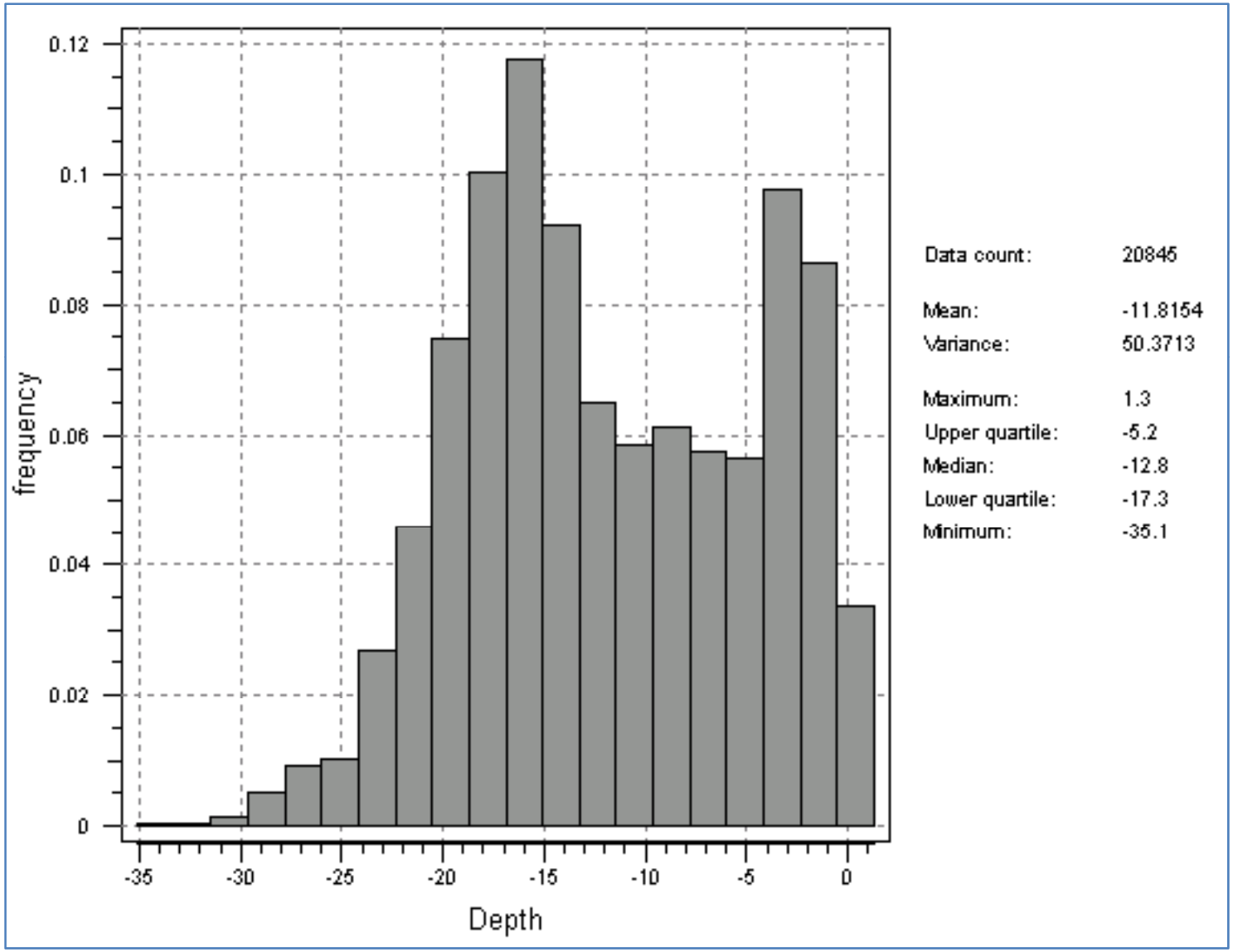


Figure 2c. Histogram and Summary Statistics for 1996 Bathymetry Survey

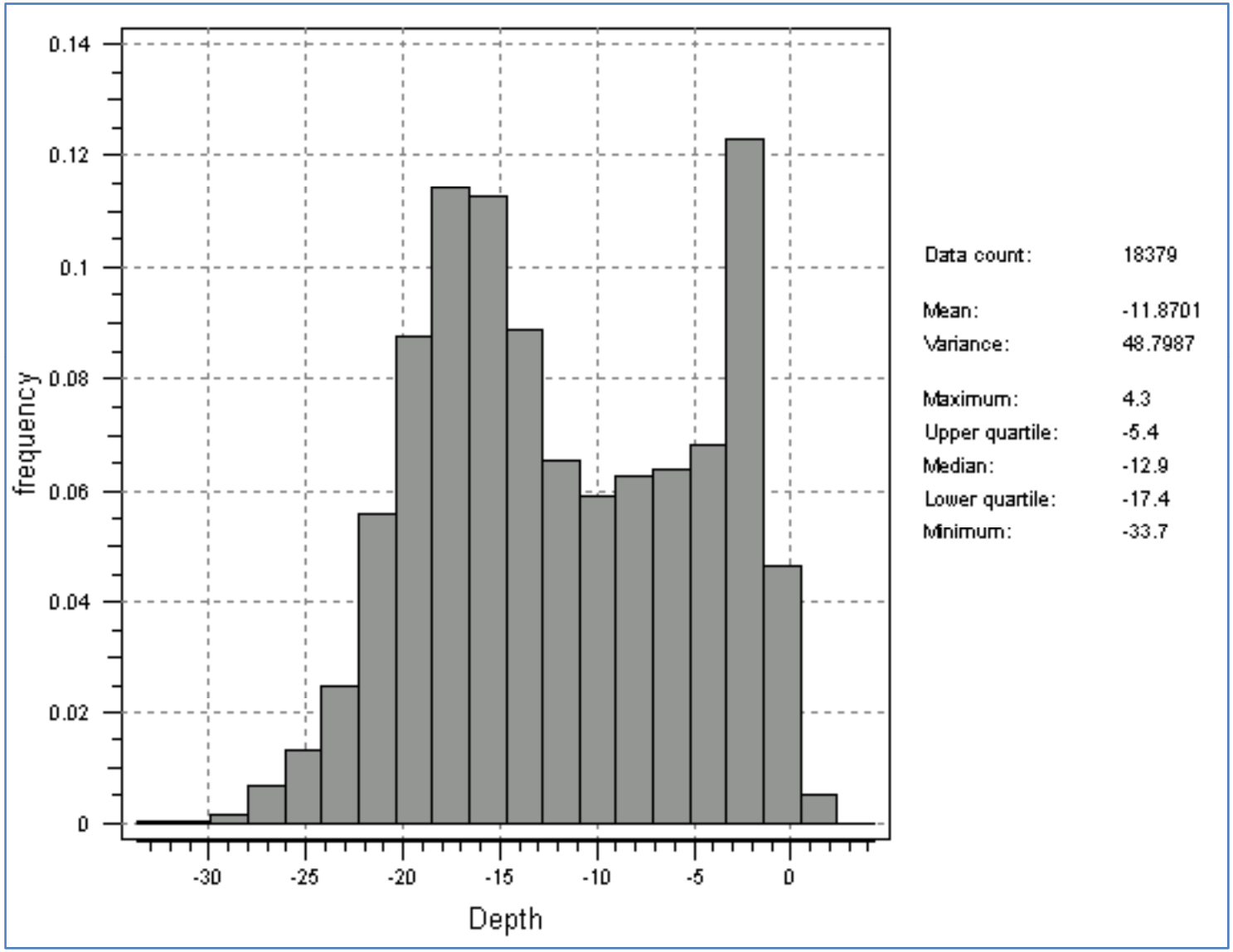


Figure 2d. Histogram and Summary Statistics for 1997 Bathymetry Survey

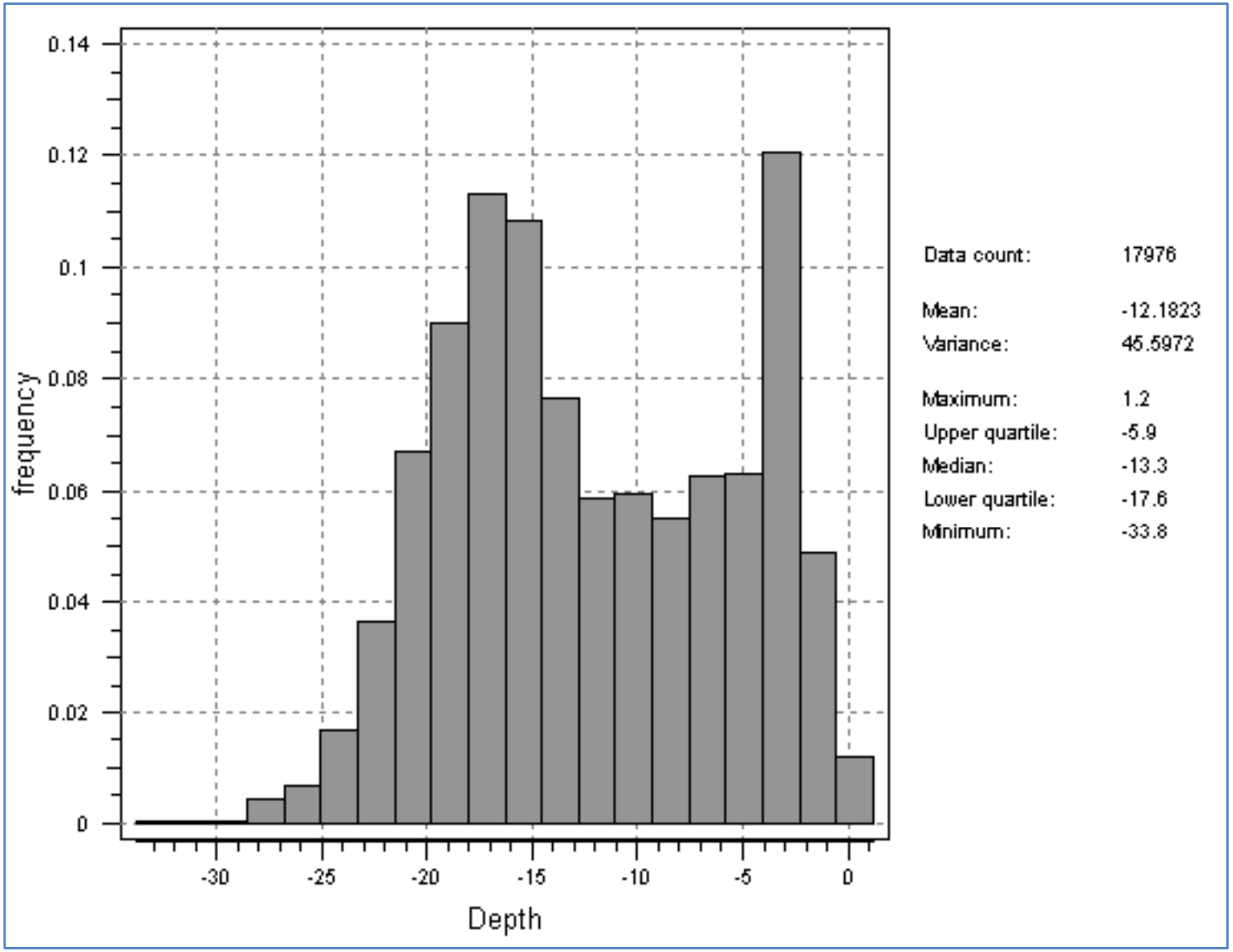


Figure 2e. Histogram and Summary Statistics for 1999 Bathymetry Survey

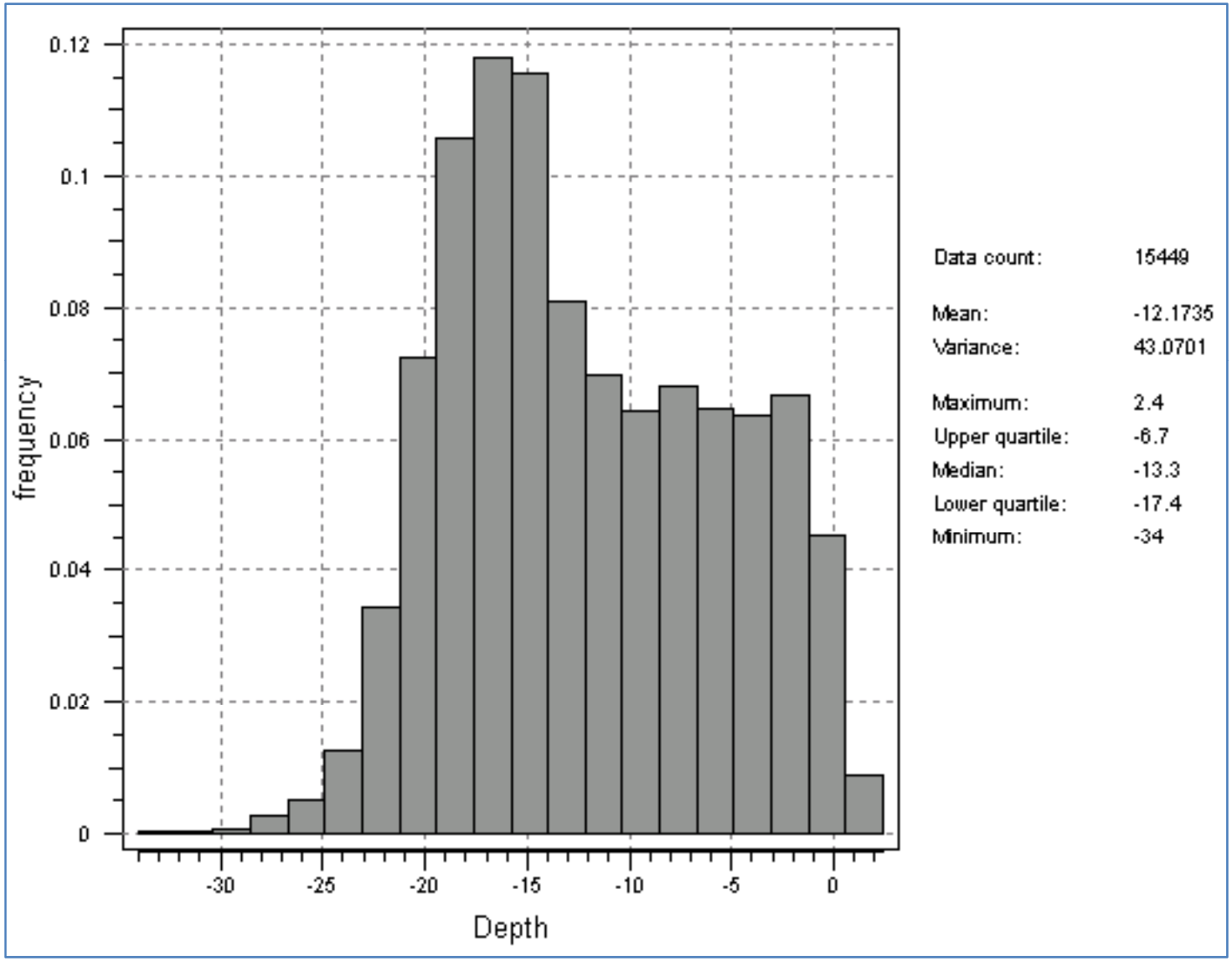


Figure 2f. Histogram and Summary Statistics for 2001 Bathymetry Survey

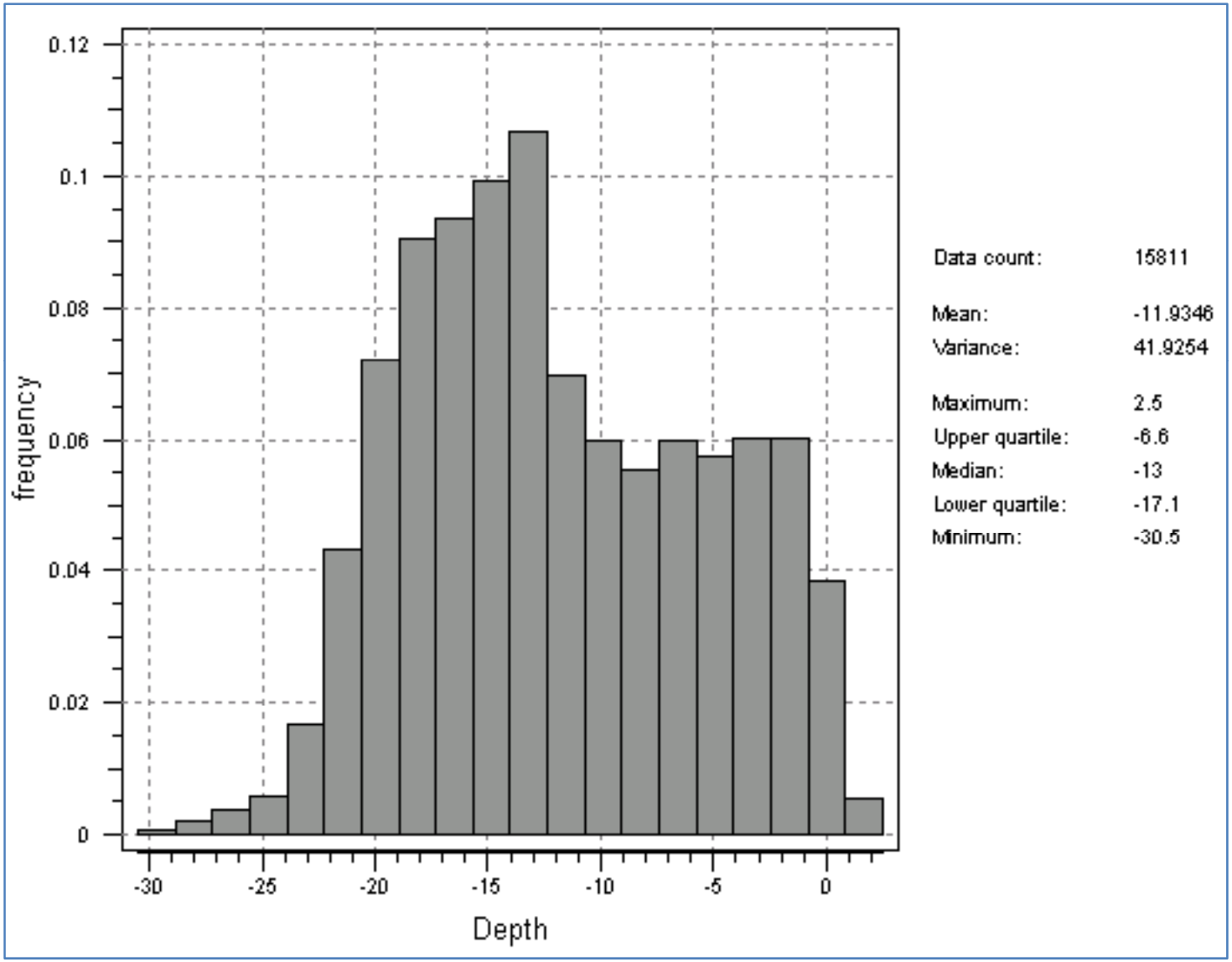
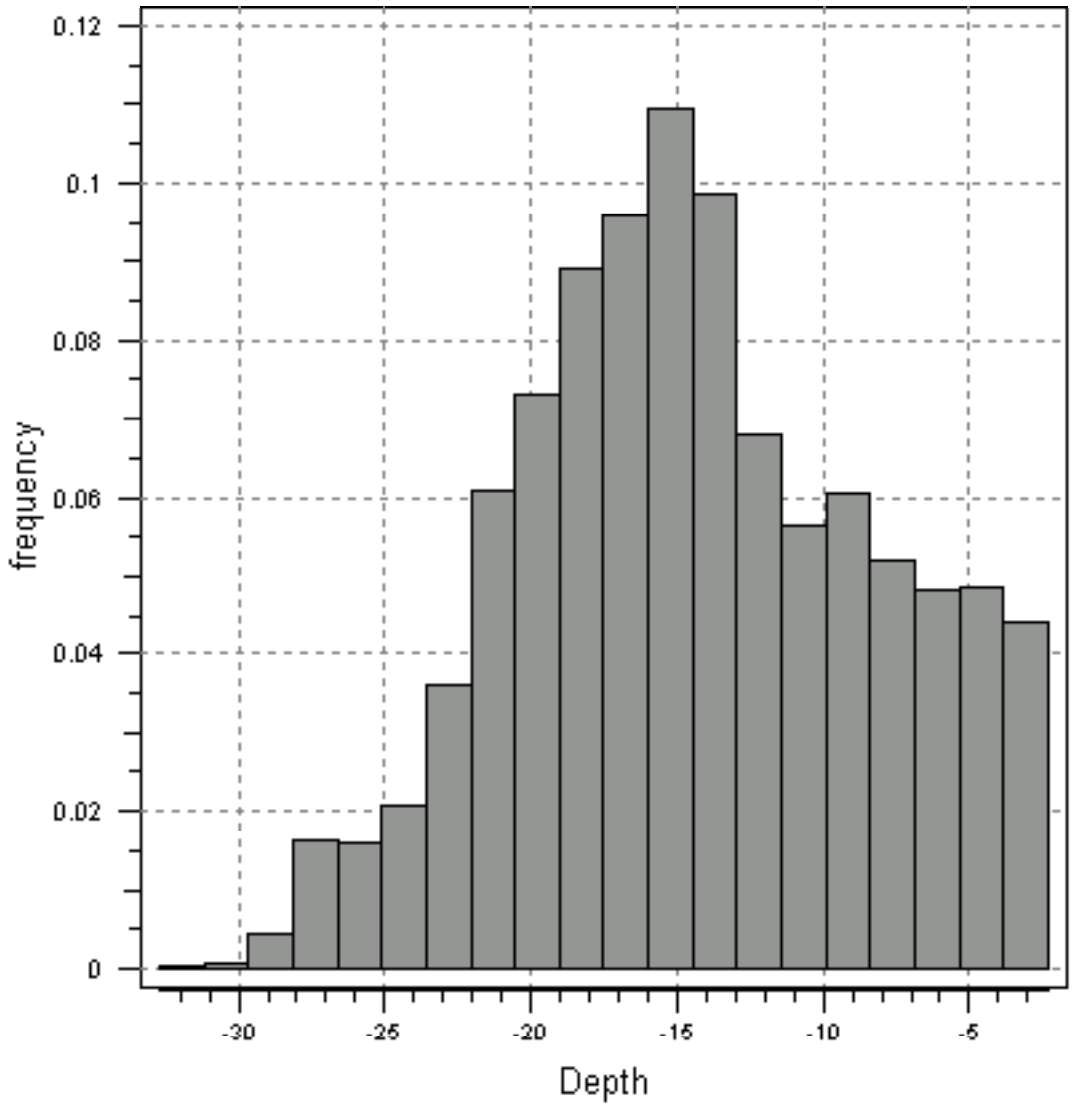
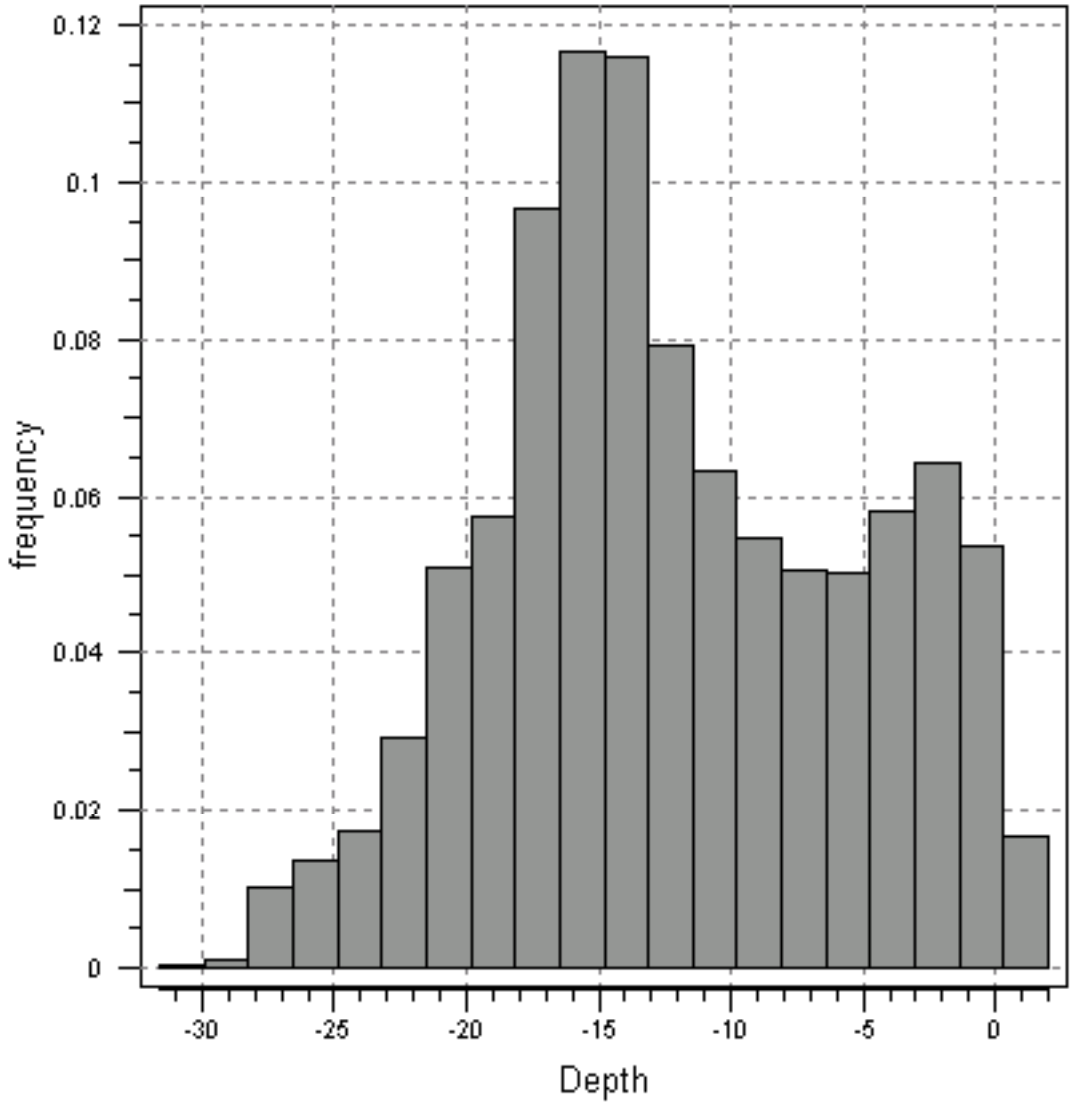


Figure 2g. Histogram and Summary Statistics for 2002 Bathymetry Survey



Data count:	18969
Mean:	-14.3821
Variance:	36.5342
Maximum:	-2.3
Upper quartile:	-9.9
Median:	-14.8
Lower quartile:	-18.7
Minimum:	-32.7

Figure 2h. Histogram and Summary Statistics for 2004 Bathymetry Survey



Data count:	24895
Mean:	-12.1407
Variance:	46.002
Maximum:	2
Upper quartile:	-6.6
Median:	-13.3
Lower quartile:	-16.9
Minimum:	-31.6

Figure 2i: Polynomial Regression Result for 1989 bathymetry

Multiple Regression Analysis for 1989 Bathymetry Survey Depth and Transformed River Coordinates
(Note: U = Cross Flow Coordinate; V = Along Flow Coordinate)

Dependent variable: Depth

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-14.62280	0.236888	-61.729	0.0000
U	2.47516	0.215729	11.474	0.0000
V	0.03028	0.002888	10.484	0.0000
U*U	0.90473	0.112363	8.052	0.0000
V*V	-0.00015	0.000008	-17.603	0.0000
U*V	-0.01657	0.000966	-17.152	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	95633.4	5	19126.7	519.73	0.0000
Residual	611305.0	16611	36.8012		
Total (Corr.)	706938.0	16616			

R-squared = 13.5278 percent
R-squared (adjusted for d.f.) = 13.5018 percent
Standard Error of Est. = 6.0664
Mean absolute error = 4.79136
Durbin-Watson statistic = 0.130868 (P=0.0000)
Lag 1 residual autocorrelation = 0.9345

Figure 2j: Polynomial Regression Result for 1995 bathymetry. Inset Figure is Semivariogram of Residuals

Multiple Regression Analysis for 1995 Bathymetry Survey Depth and Transformed River Coordinates
(Note: U = Cross Flow Coordinate; V = Along Flow Coordinate)

Dependent variable: Depth

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-6.01978	0.474852	-12.6772	0.0000
U	-1.09495	0.228131	-4.7997	0.0000
V	-0.08985	0.005290	-16.9857	0.0000
U*U	5.32106	0.091953	57.8671	0.0000
V*V	0.00022	0.000014	15.9645	0.0000
U*V	-0.00474	0.001005	-4.7121	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	181699.0	5	36339.8	872.21	0.0000
Residual	868240.0	20839	41.6642		
Total (Corr.)	1.04994E6	20844			

R-squared = 17.3057 percent
R-squared (adjusted for d.f.) = 17.2858 percent
Standard Error of Est. = 6.45478
Mean absolute error = 5.33981
Durbin-Watson statistic = 0.0296894 (P=0.0000)
Lag 1 residual autocorrelation = 0.985152

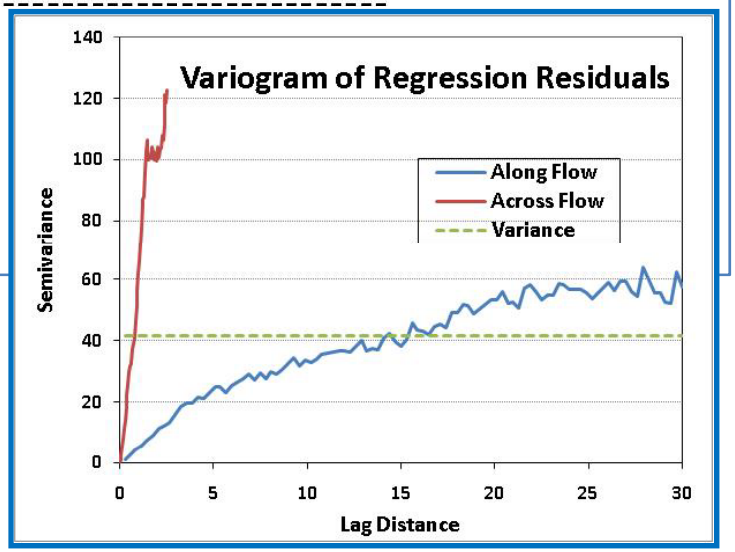


Figure 2k: Polynomial Regression Result for 1996 bathymetry

Multiple Regression Analysis for 1996 Bathymetry Survey Depth and Transformed River Coordinates
Note: U = Cross Flow Coordinate; V = Along Flow Coordinate)

Dependent variable: Depth

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-16.90500	0.899662	-18.7904	0.0000
U	-0.36220	0.318288	-1.1380	0.2551
V	0.00854	0.008793	0.9716	0.3313
U*U	5.16838	0.095408	54.1717	0.0000
V*V	0.00001	0.000021	0.4037	0.6864
U*V	-0.00683	0.001346	-5.0771	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	149683.0	5	29936.6	736.17	0.0000
Residual	747140.0	18373	40.6651		
Total (Corr.)	896823.0	18378			

R-squared = 16.6904 percent
R-squared (adjusted for d.f.) = 16.6677 percent
Standard Error of Est. = 6.37692
Mean absolute error = 5.26645
Durbin-Watson statistic = 1.95342 (P=0.0008)
Lag 1 residual autocorrelation = 0.0232867

Figure 2l: Polynomial Regression Result for 1997 bathymetry

Multiple Regression Analysis for 1997 Bathymetry Survey Depth and Transformed River Coordinates
(Note: U = Cross Flow Coordinate; V = Along Flow Coordinate)

Dependent variable: Depth

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-15.43800	0.889624	-17.3534	0.0000
U	0.41462	0.327394	1.2664	0.2054
V	-0.00962	0.008684	-1.1077	0.2680
U*U	4.62971	0.098426	47.0377	0.0000
V*V	0.00006	0.000020	2.8574	0.0043
U*V	-0.00960	0.001377	-6.9715	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	122616.0	5	24523.2	632.26	0.0000
Residual	696994.0	17970	38.7865		
Total (Corr.)	819610.0	17975			

R-squared = 14.9603 percent
R-squared (adjusted for d.f.) = 14.9366 percent
Standard Error of Est. = 6.22788
Mean absolute error = 5.1599
Durbin-Watson statistic = 1.46572 (P=0.0000)
Lag 1 residual autocorrelation = 0.267112

Figure 2m: Polynomial Regression Result for 1999 bathymetry

Multiple Regression Analysis for 1999 Bathymetry Survey Depth and Transformed River Coordinates
(Note: U = Cross Flow Coordinate; V = Along Flow Coordinate)

Dependent variable: Depth

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-23.76950	0.87765	-27.0832	0.0000
U	-9.85349	0.31098	-31.6848	0.0000
V	0.08111	0.00884	9.1712	0.0000
U*U	7.05628	0.09114	77.4202	0.0000
V*V	-0.00019	0.00002	-8.9664	0.0000
U*V	0.04073	0.00137	29.702	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	193813.0	5	38762.6	1269.50	0.0000
Residual	471533.0	15443	30.5338		
Total (Corr.)	665346.0	15448			

R-squared = 29.1296 percent
R-squared (adjusted for d.f.) = 29.1067 percent
Standard Error of Est. = 5.52574
Mean absolute error = 4.52886
Durbin-Watson statistic = 1.59393 (P=0.0000)
Lag 1 residual autocorrelation = 0.202988

Figure 2n: Polynomial Regression Result for 2001 bathymetry

Multiple Regression Analysis for 2001 Bathymetry Survey Depth and Transformed River Coordinates
(Note: U = Cross Flow Coordinate; V = Along Flow Coordinate)

Dependent variable: Depth

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-26.8546	0.86168	-31.1653	0.0000
U	-10.7103	0.30563	-35.0432	0.0000
V	0.1128	0.00868	12.9854	0.0000
U*U	6.7680	0.08866	76.3319	0.0000
V*V	-0.0003	0.00002	-12.4402	0.0000
U*V	0.0447	0.00134	33.2957	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	194485.0	5	38896.9	1312.61	0.0000
Residual	468355.0	15805	29.6334		
Total (Corr.)	662840.0	15810			

R-squared = 29.3411 percent
R-squared (adjusted for d.f.) = 29.3188 percent
Standard Error of Est. = 5.44365
Mean absolute error = 4.47725
Durbin-Watson statistic = 1.59144 (P=0.0000)
Lag 1 residual autocorrelation = 0.204263

Figure 2o: Polynomial Regression Result for 2002 bathymetry

Multiple Regression Analysis for 2002 Bathymetry Survey Depth and Transformed River Coordinates
(Note: U = Cross Flow Coordinate; V = Along Flow Coordinate)

Dependent variable: Depth

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-15.1029	0.481731	-31.3513	0.0000
U	2.3383	0.233236	10.0253	0.0000
V	0.0291	0.005260	5.5312	0.0000
U*U	1.1821	0.094716	12.4807	0.0000
V*V	-0.0001	0.000013	-8.4951	0.0000
U*V	-0.0176	0.001022	-17.2185	0.0000

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	65099.7	5	13019.9	393.22	0.0000
Residual	627880.0	18963	33.1108		
Total (Corr.)	692980.0	18968			

R-squared = 9.39416 percent
R-squared (adjusted for d.f.) = 9.37027 percent
Standard Error of Est. = 5.7542
Mean absolute error = 4.7017
Durbin-Watson statistic = 0.0507036 (P=0.0000)
Lag 1 residual autocorrelation = 0.974557

Figure 2p: Polynomial Regression Result for 2004 bathymetry

Multiple Regression Analysis for 2004 Bathymetry Survey Depth and Transformed River Coordinates
(Note: U = Cross Flow Coordinate; V = Along Flow Coordinate)

Dependent variable: Depth

Parameter	Estimate	Standard Error	T Statistic	P-Value
CONSTANT	-13.6587	0.164017	-83.2761	0.0000
U	-1.95215	0.136465	-14.3052	0.0000
V	0.03363	0.002168	15.5093	0.0000
U*U	4.23460	0.076119	55.6312	0.0000
V*V	-0.00016	0.000007	-24.7688	0.0000
U*V	-0.00185	0.000639	-2.8930	0.0038

Analysis of Variance

Source	Sum of Squares	Df	Mean Square	F-Ratio	P-Value
Model	207735.0	5	41546.9	1103.07	0.0000
Residual	937440.0	24889	37.6648		
Total (Corr.)	1.14517E6	24894			

R-squared = 18.14 percent
R-squared (adjusted for d.f.) = 18.1236 percent
Standard Error of Est. = 6.13717
Mean absolute error = 5.03368
Durbin-Watson statistic = 0.0500437 (P=0.0000)
Lag 1 residual autocorrelation = 0.974926

Figure 3: Estimated trend surface for 1995 bathymetry elevations. Note that the long- and cross-flow coordinates are not to scale or proportion.

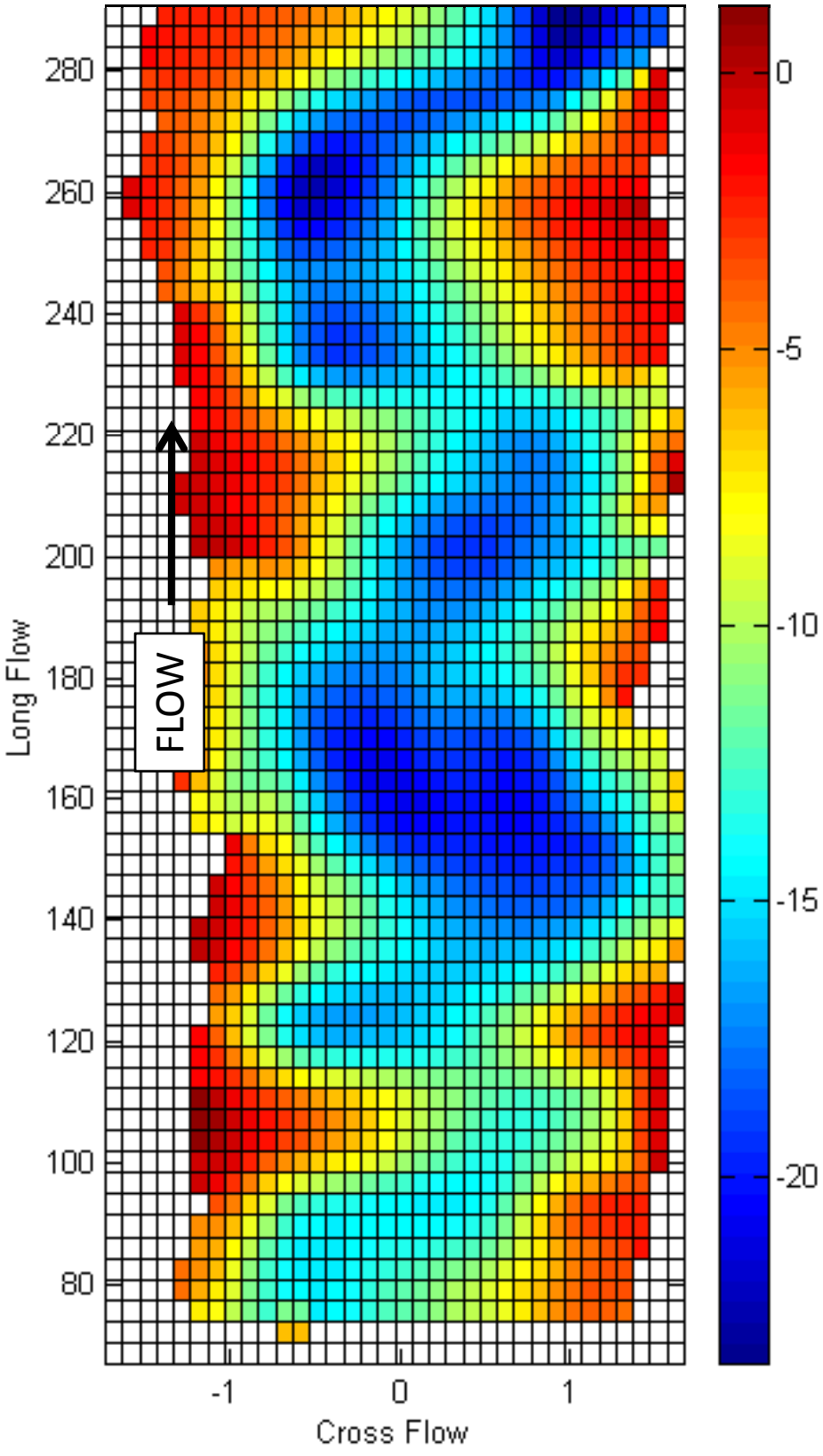
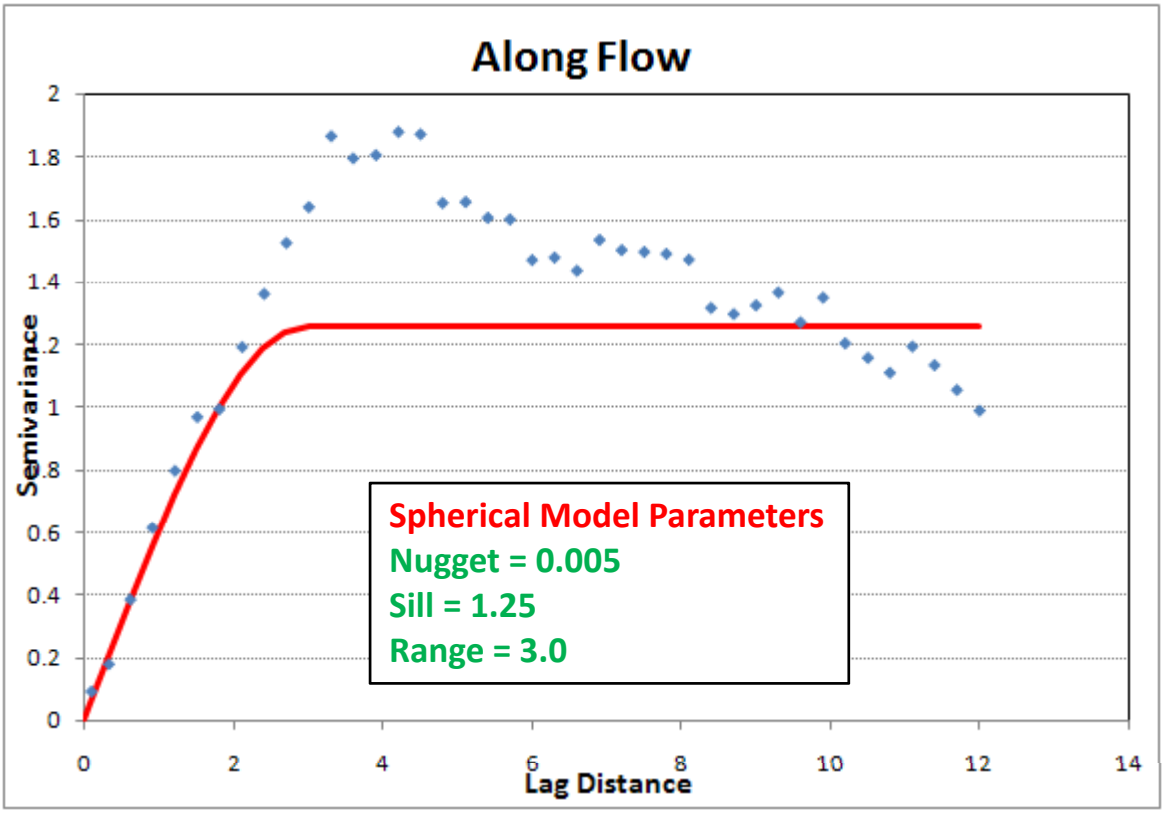


Figure 4a: Directional Semivariograms and Fitted Models for Normal Scores Transformed Residuals of Detrended 1989 Bathymetry



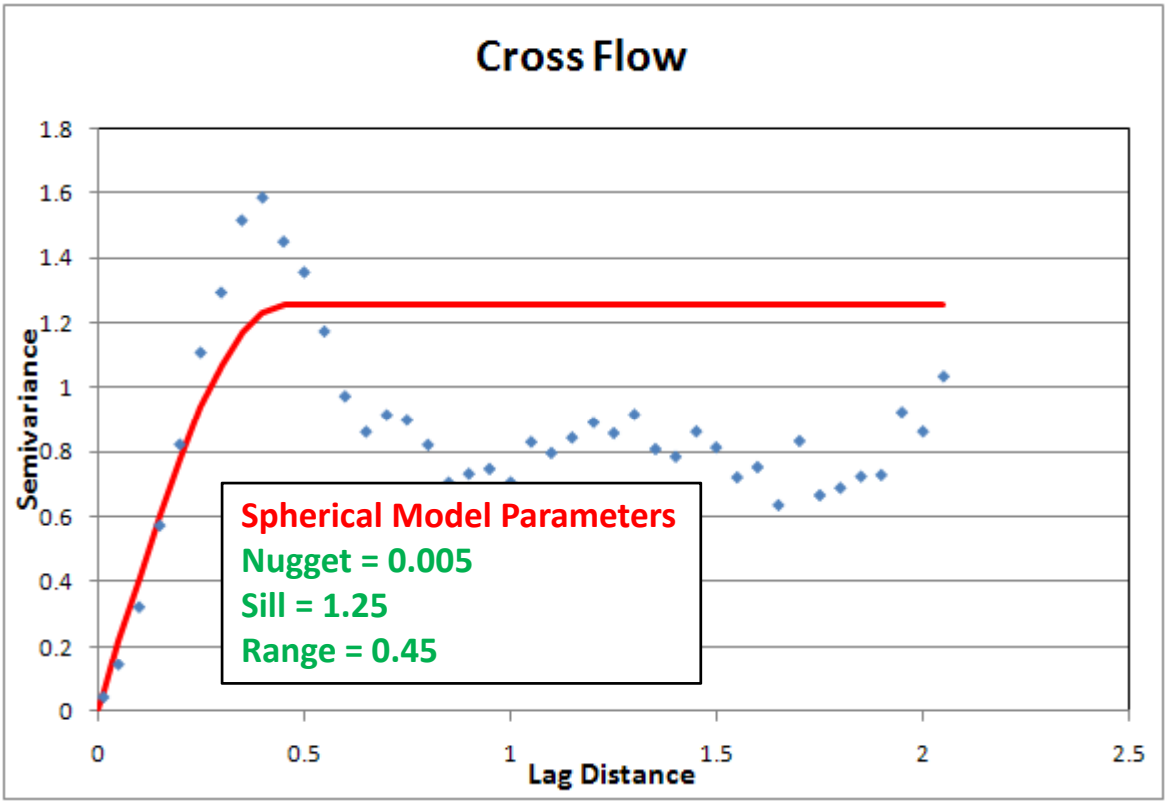
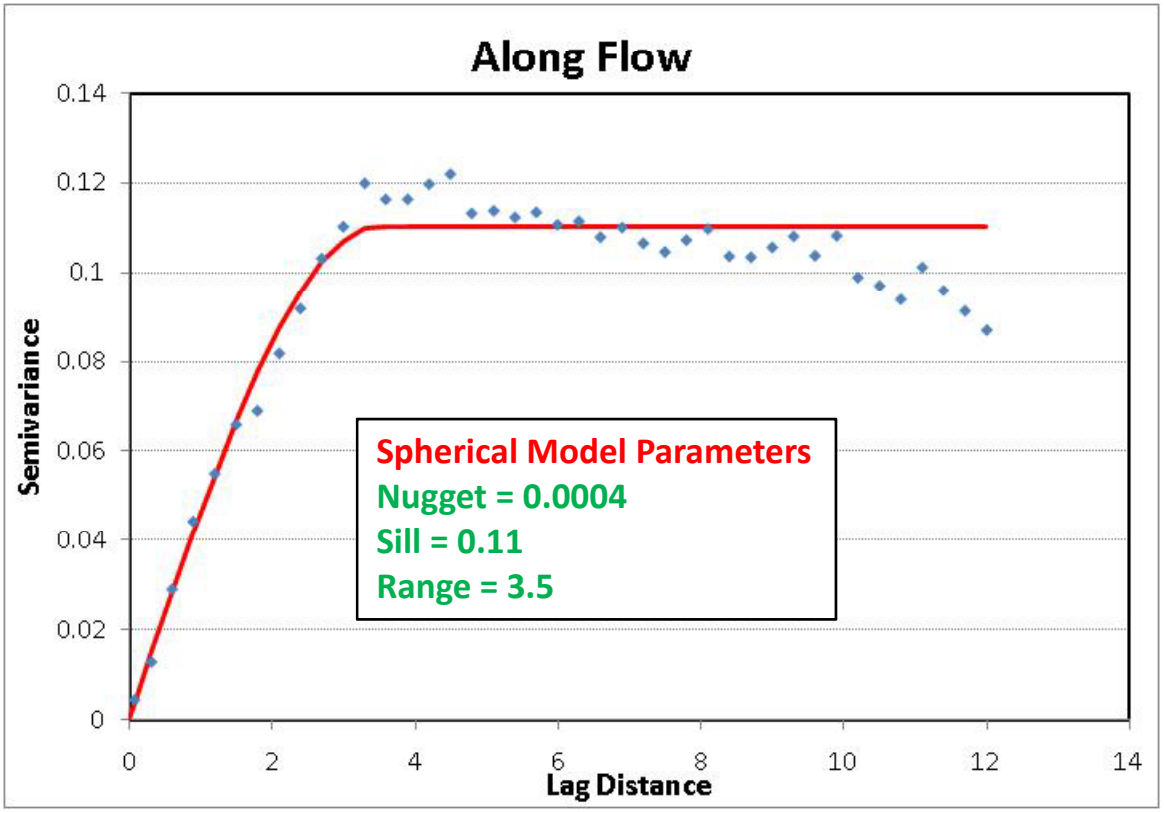


Figure 4b: Directional Semivariograms and Fitted Models for Uniform Scores Transformed Residuals of Detrended 1989 Bathymetry



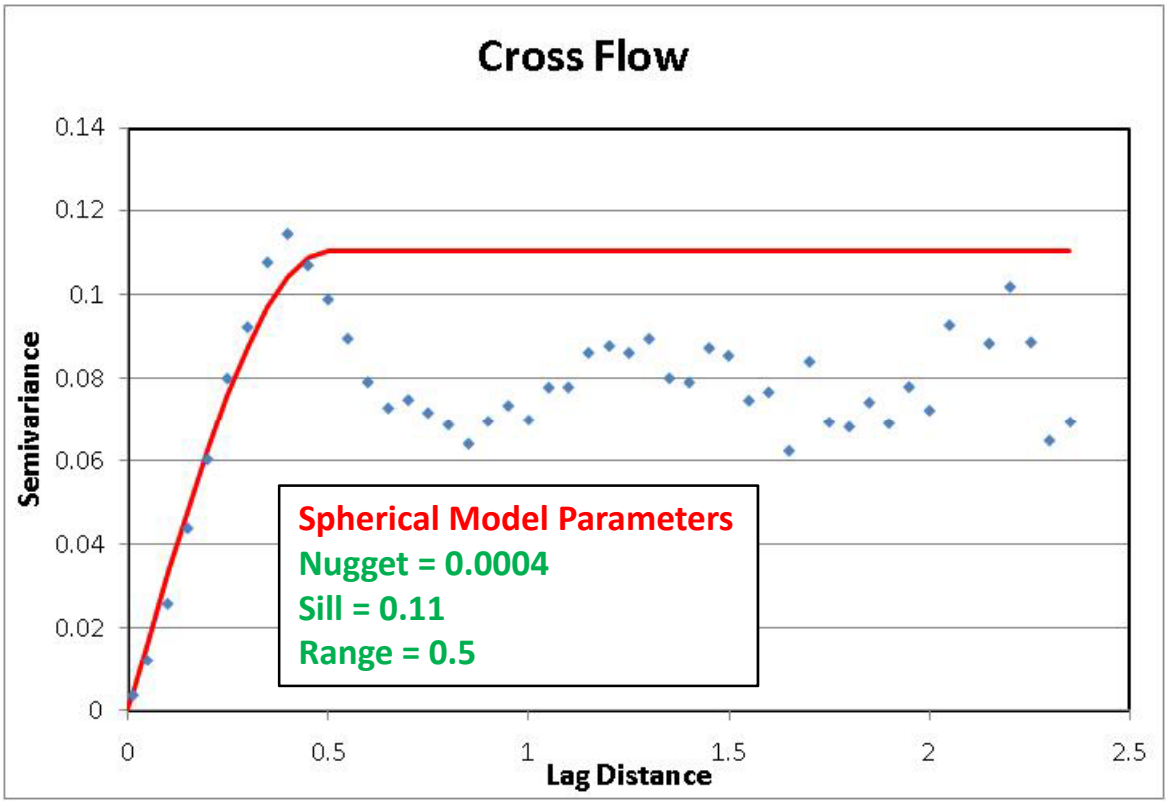
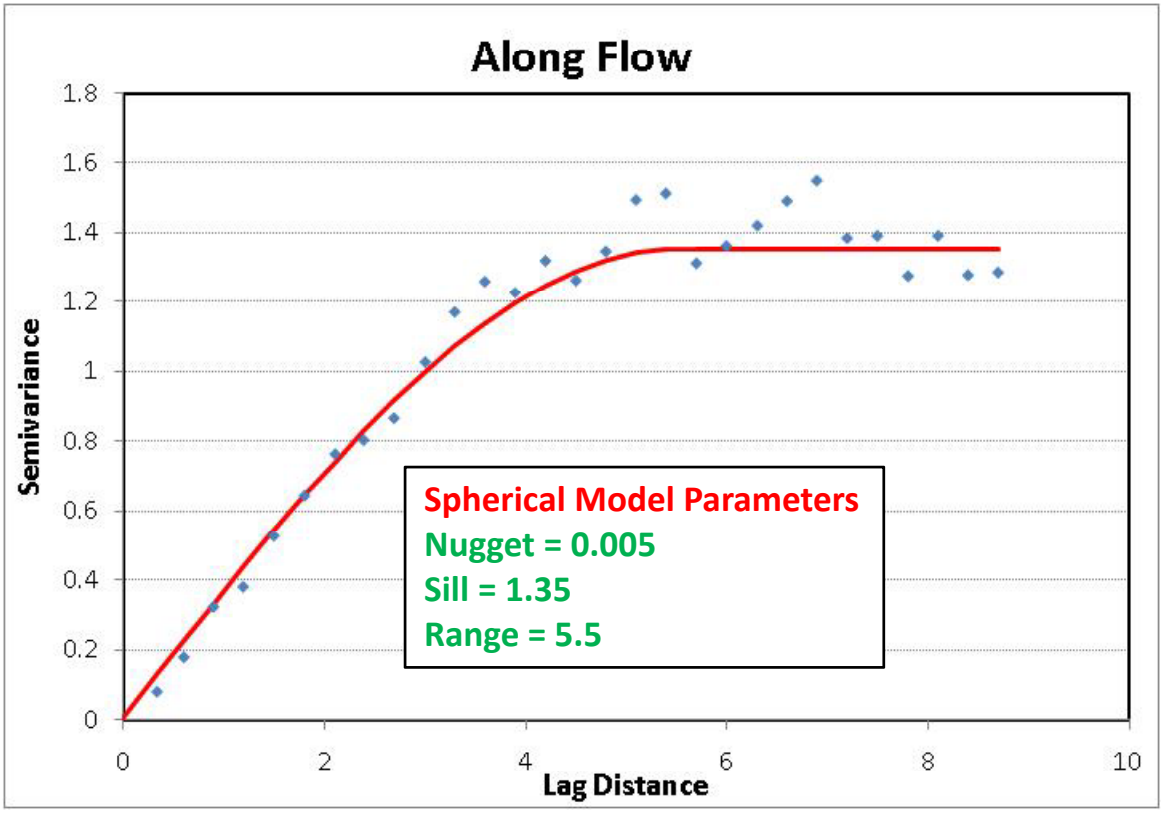


Figure 4c: Directional Semivariograms and Fitted Models for Normal Scores Transformed Residuals of Detrended 1995 Bathymetry



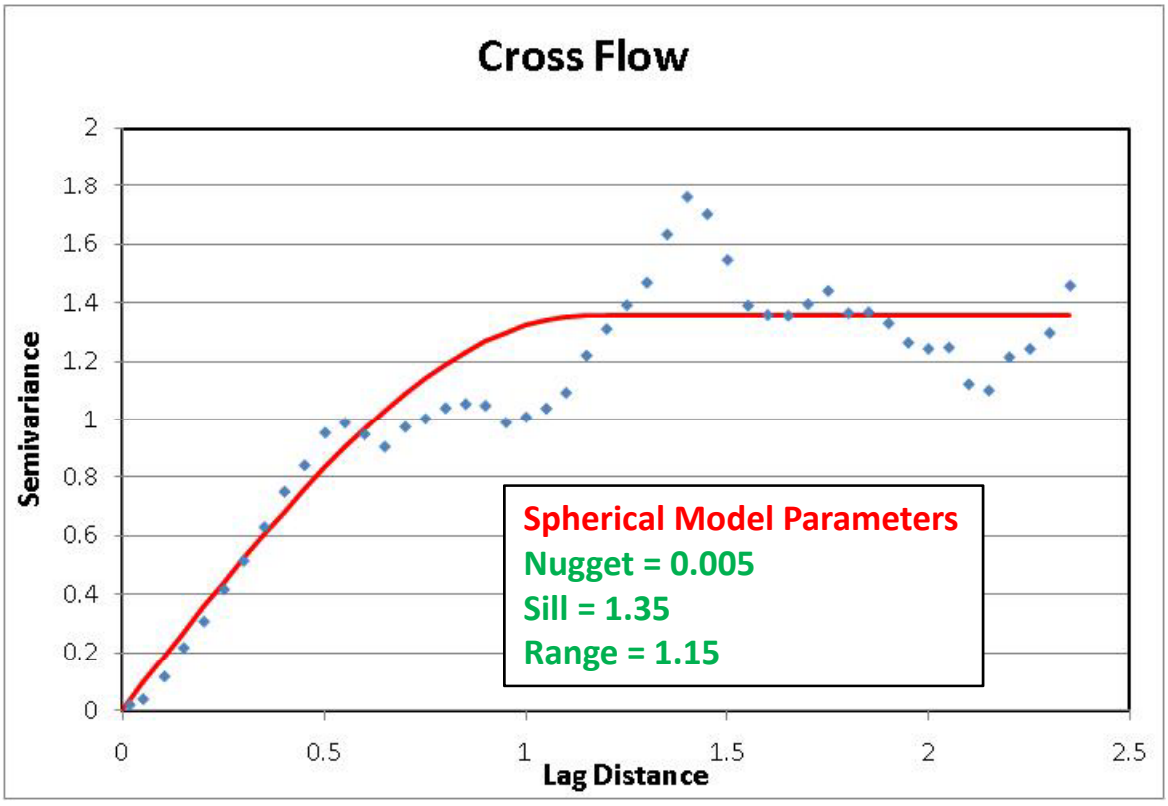
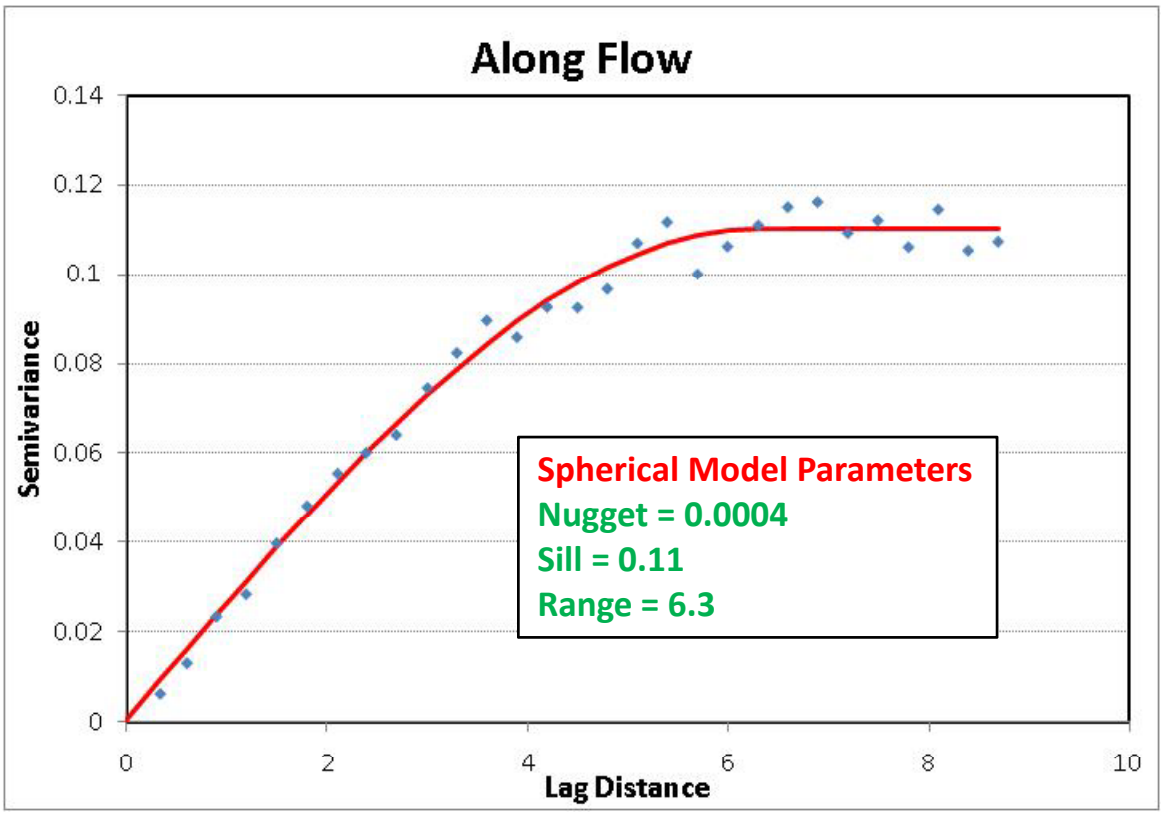


Figure 4d: Directional Semivariograms and Fitted Models for Uniform Scores Transformed Residuals of Detrended 1995 Bathymetry



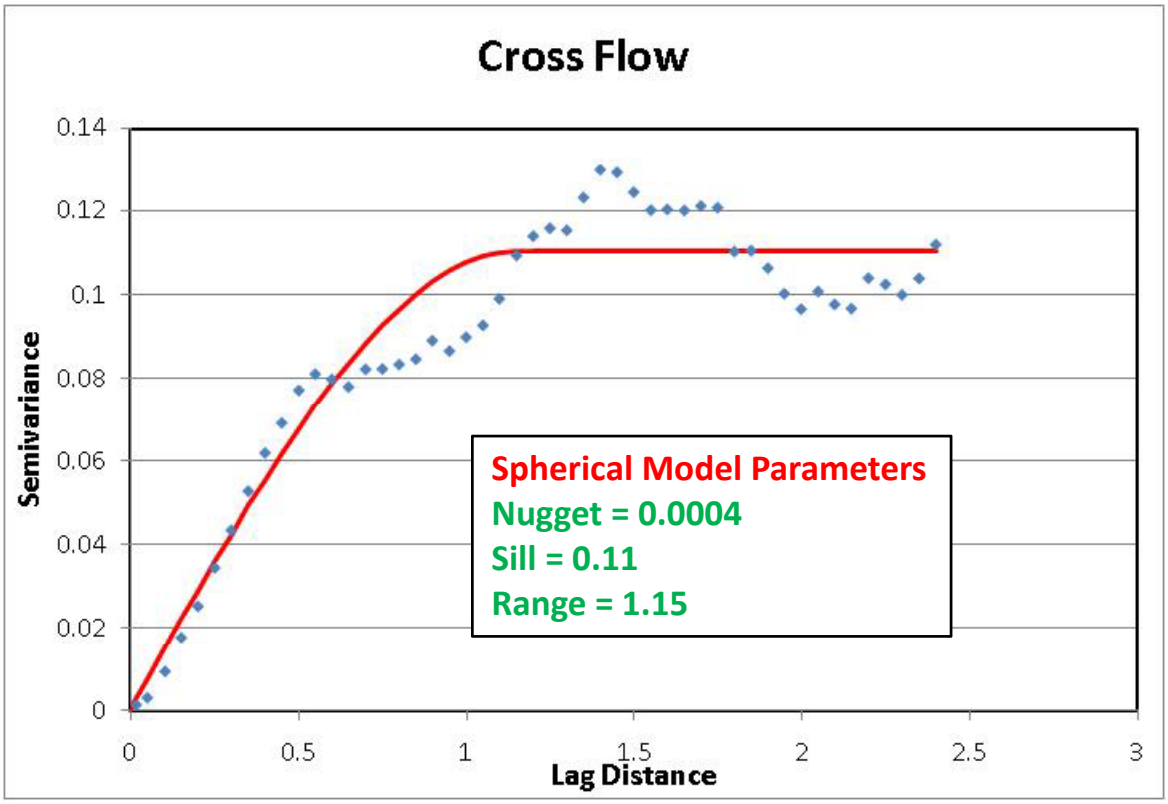
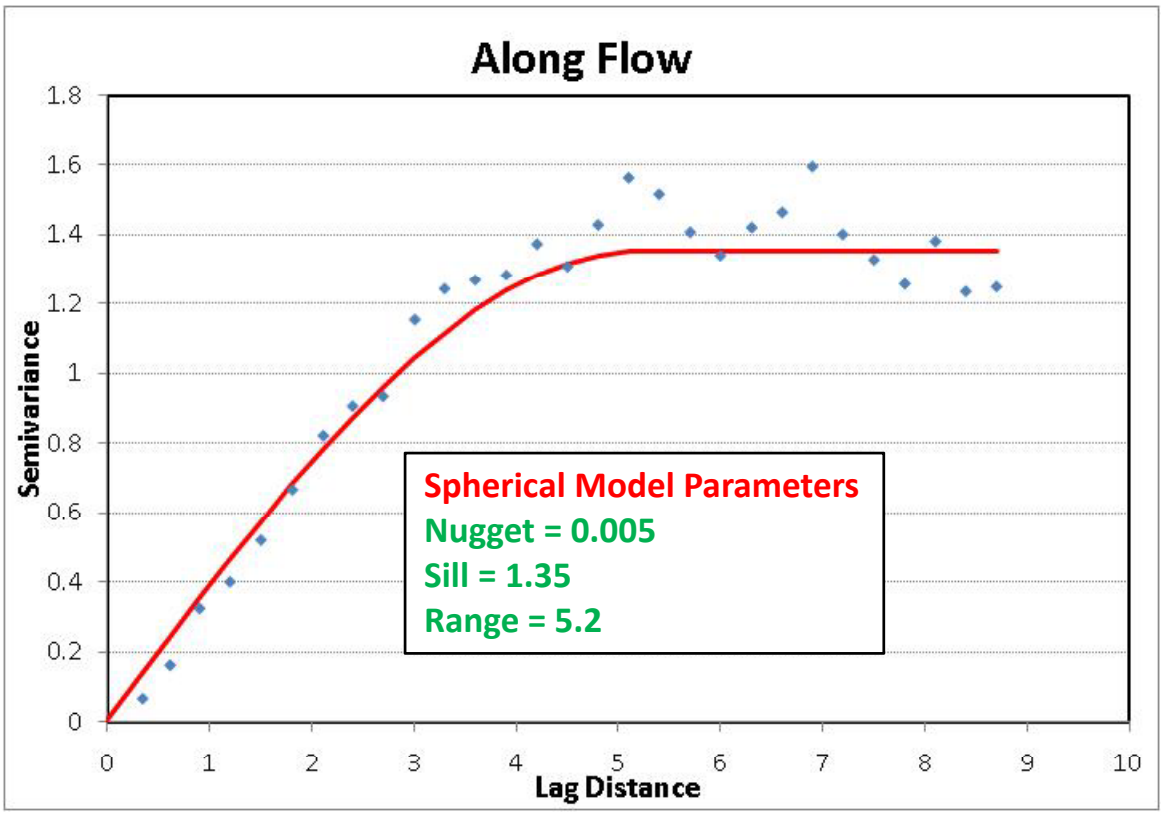


Figure 4e: Directional Semivariograms and Fitted Models for Normal Scores Transformed Residuals of Detrended 1996 Bathymetry



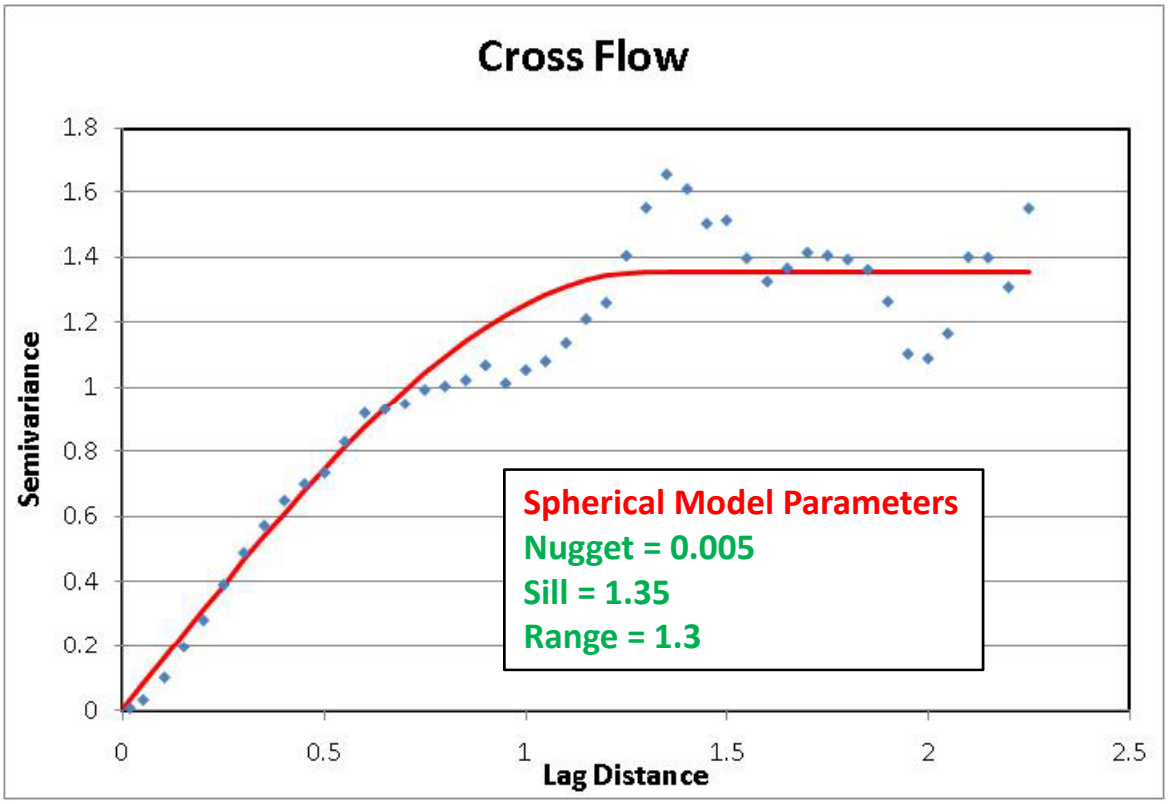
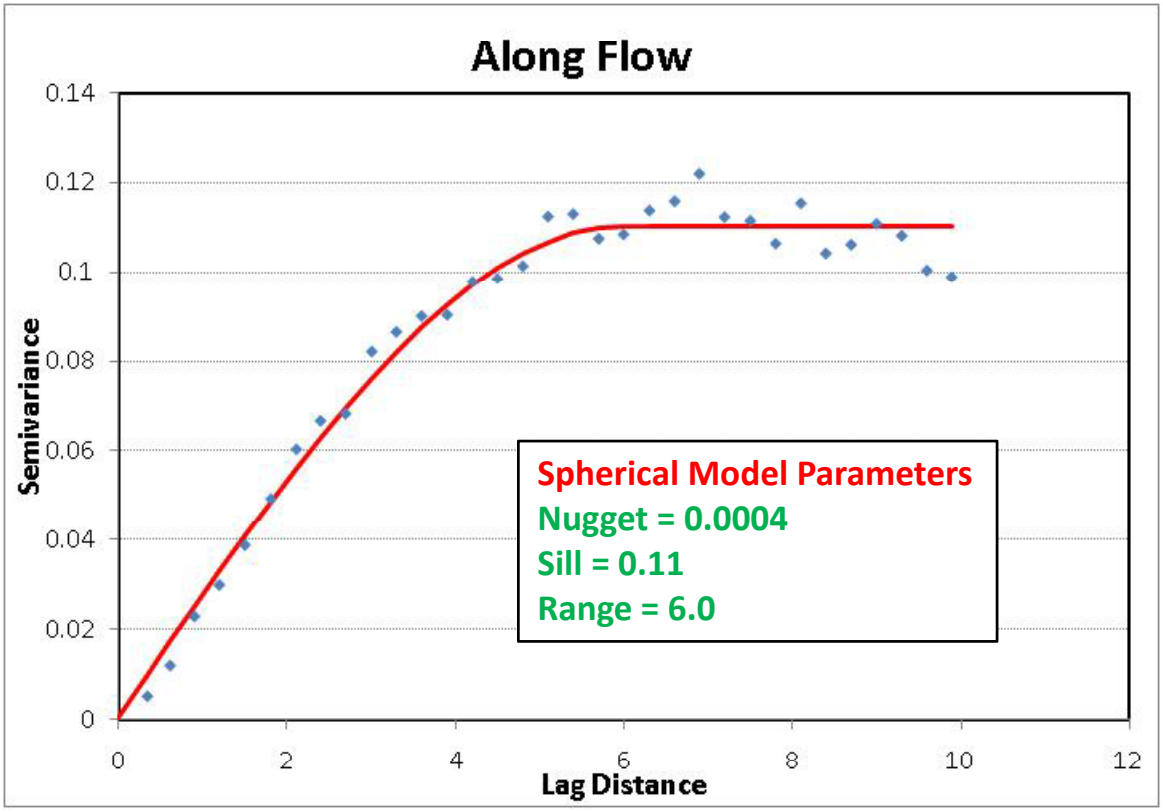


Figure 4f: Directional Semivariograms and Fitted Models for Uniform Scores Transformed Residuals of Detrended 1996 Bathymetry



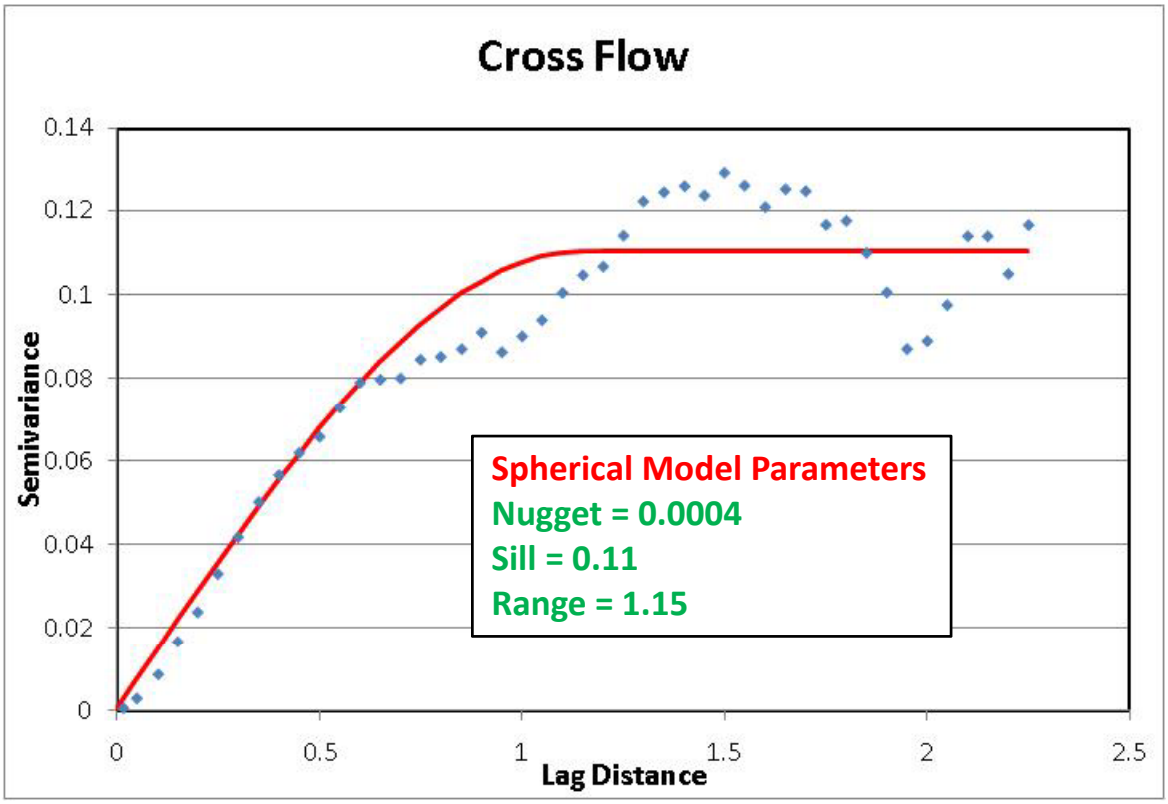
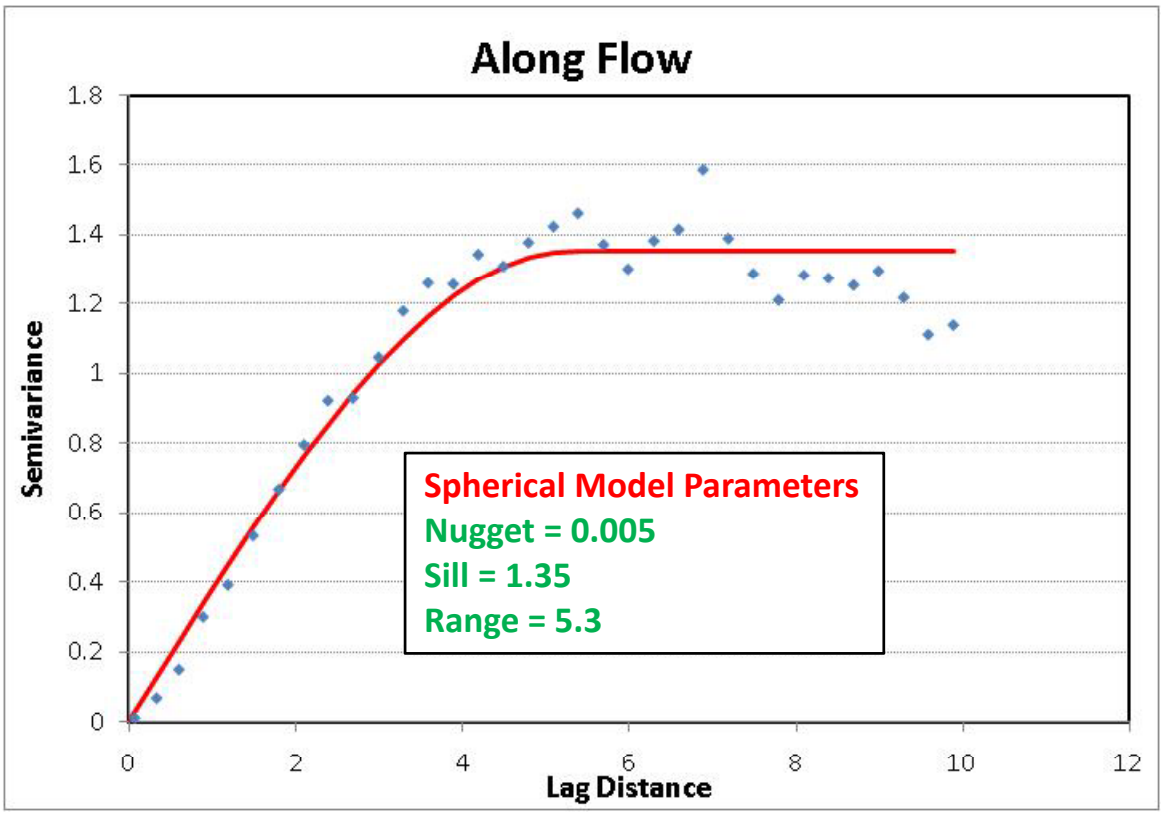


Figure 4g: Directional Semivariograms and Fitted Models for Normal Scores Transformed Residuals of Detrended 1997 Bathymetry



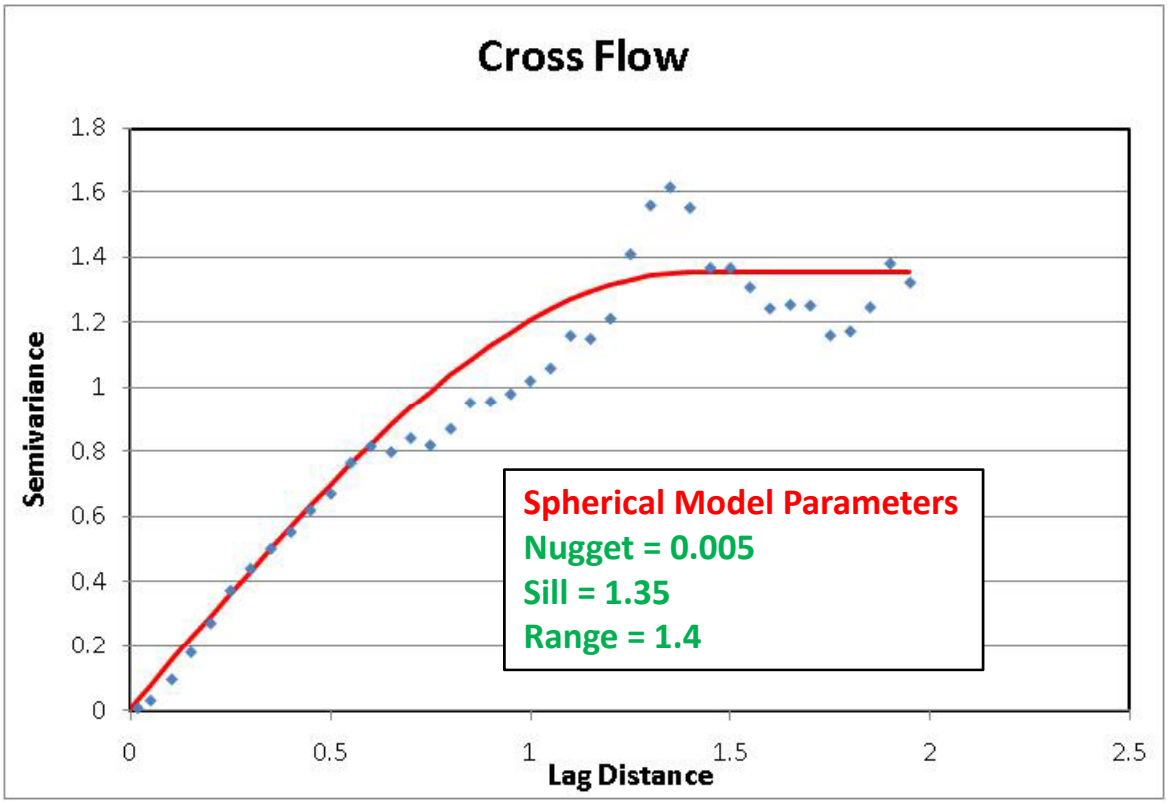
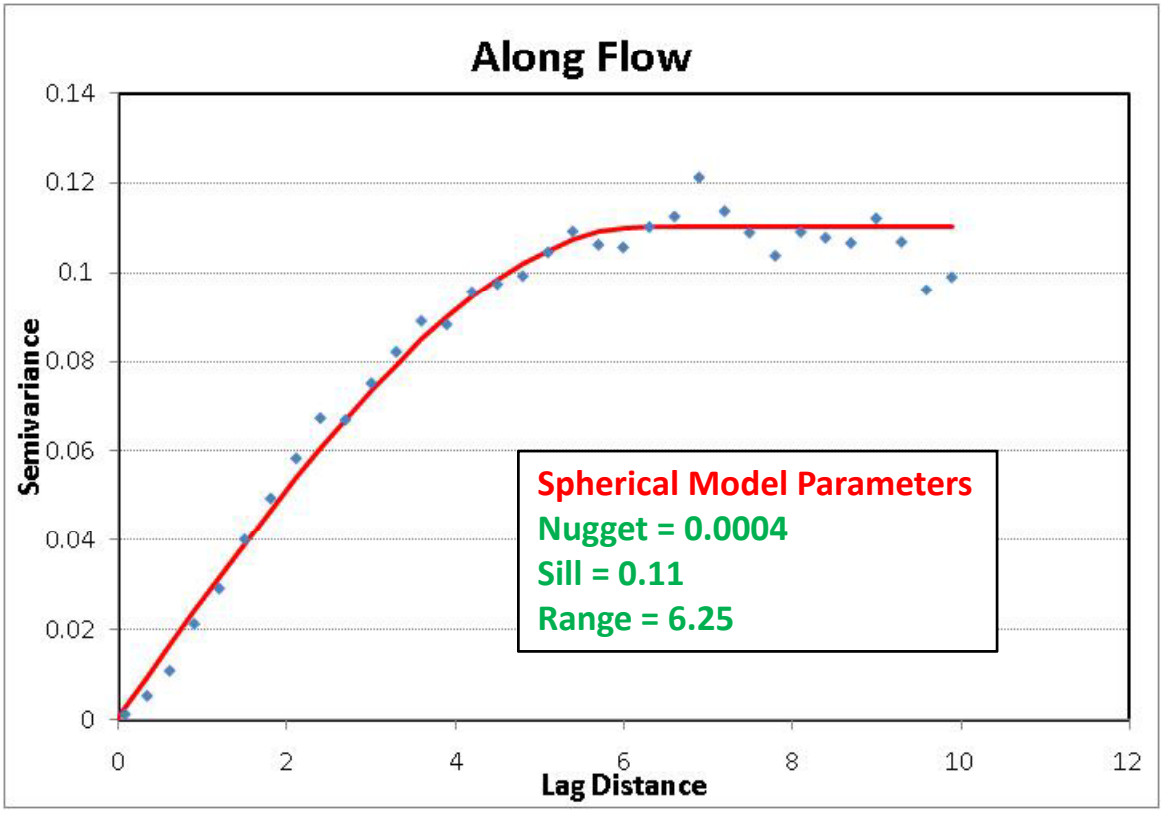


Figure 4h: Directional Semivariograms and Fitted Models for Uniform Scores Transformed Residuals of Detrended 1997 Bathymetry



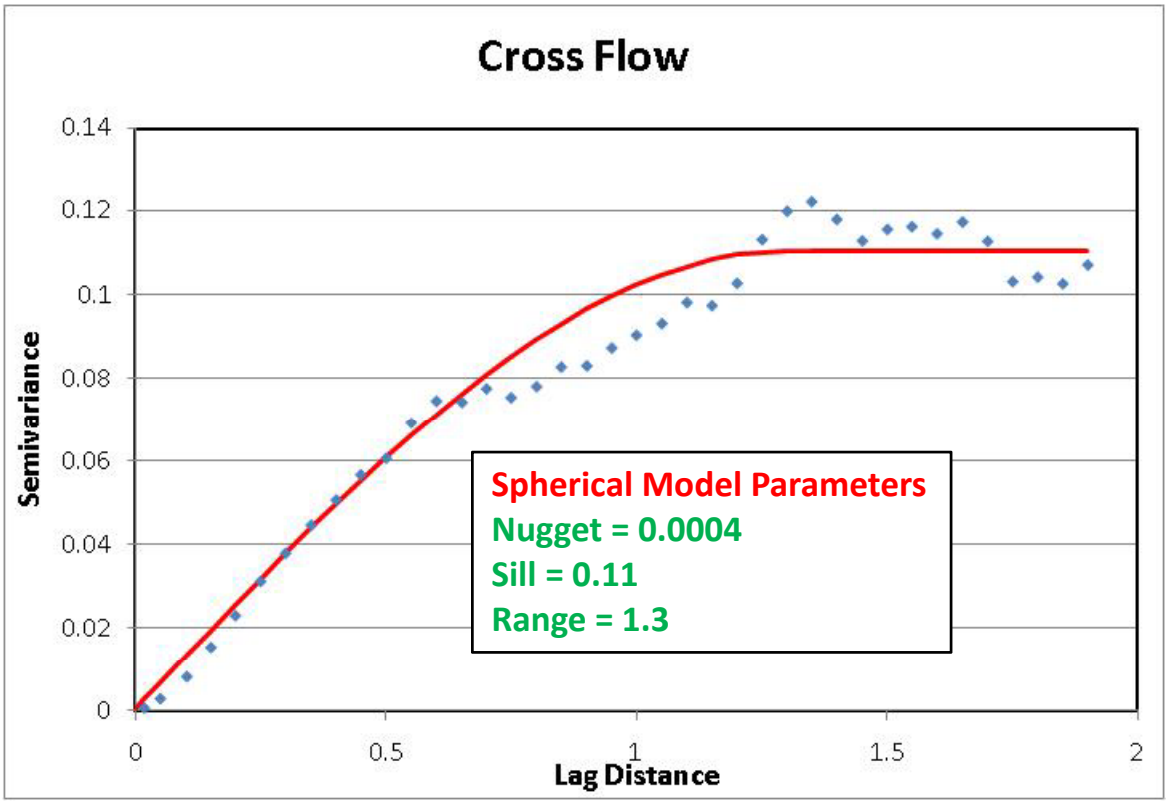
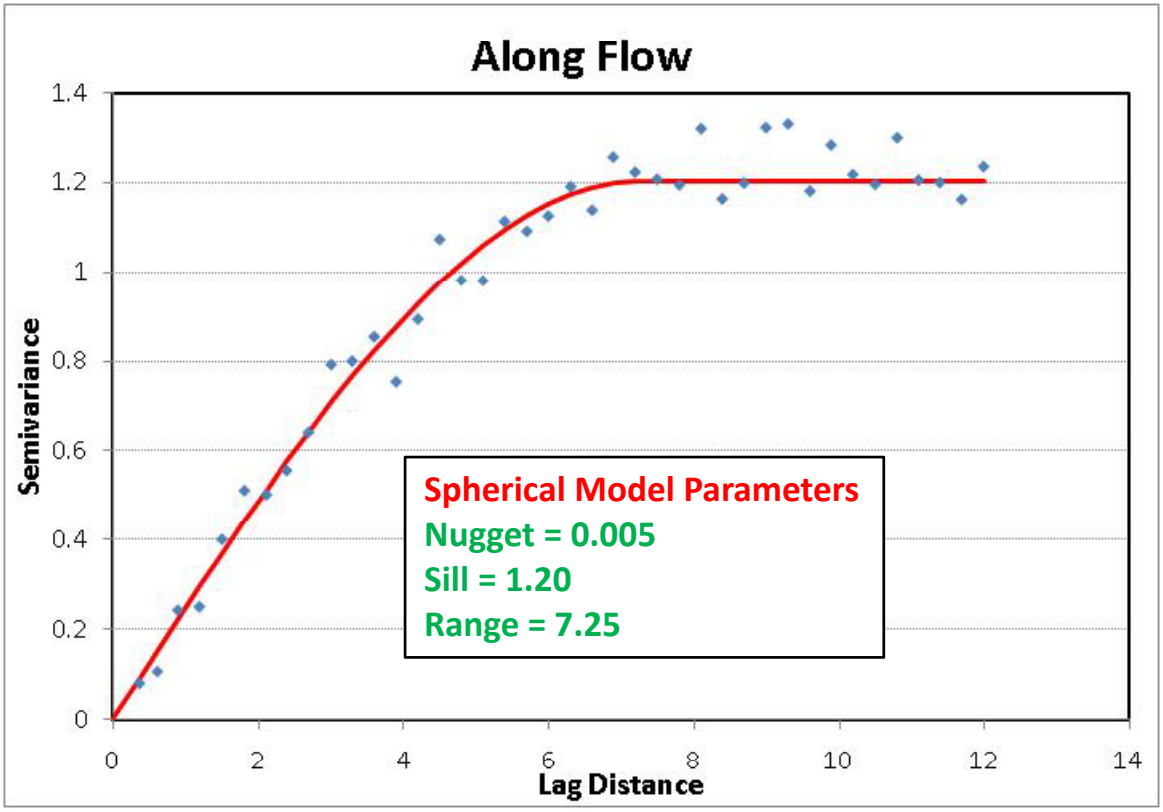


Figure 4i: Directional Semivariograms and Fitted Models for Normal Scores Transformed Residuals of Detrended 1999 Bathymetry



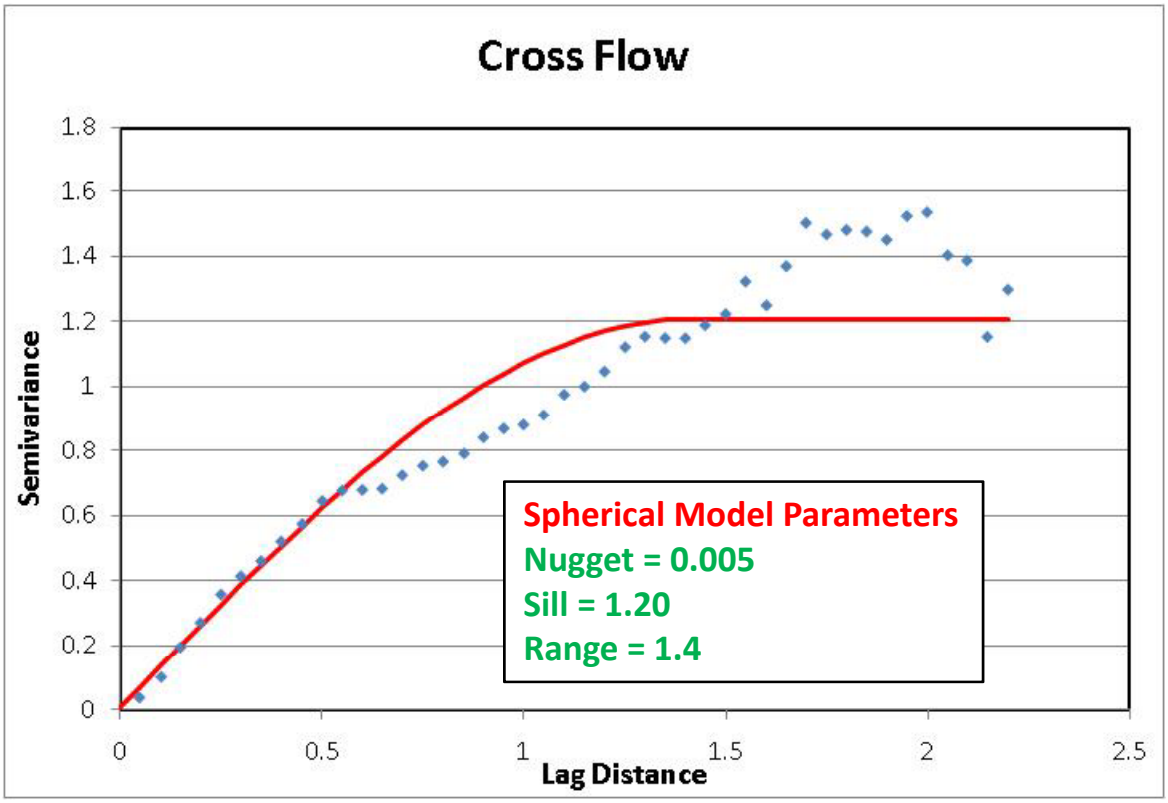
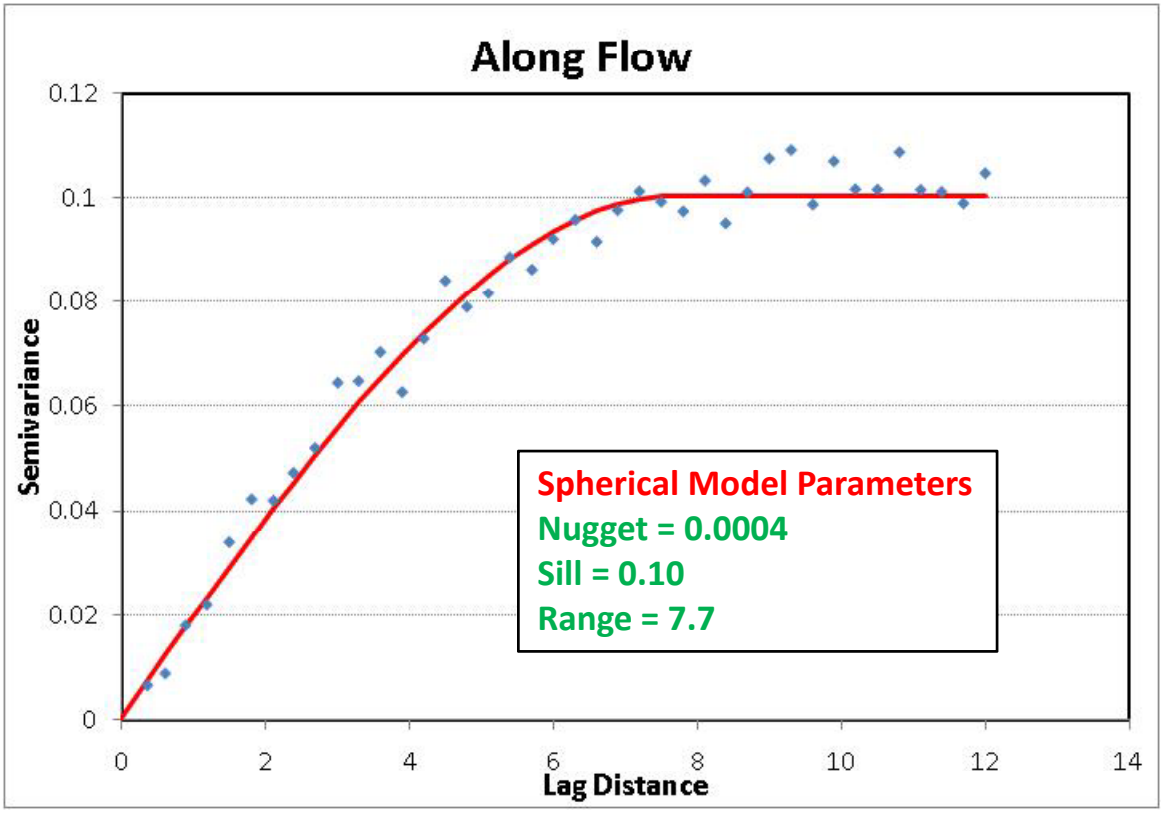


Figure 4j: Directional Semivariograms and Fitted Models for Uniform Scores Transformed Residuals of Detrended 1999 Bathymetry



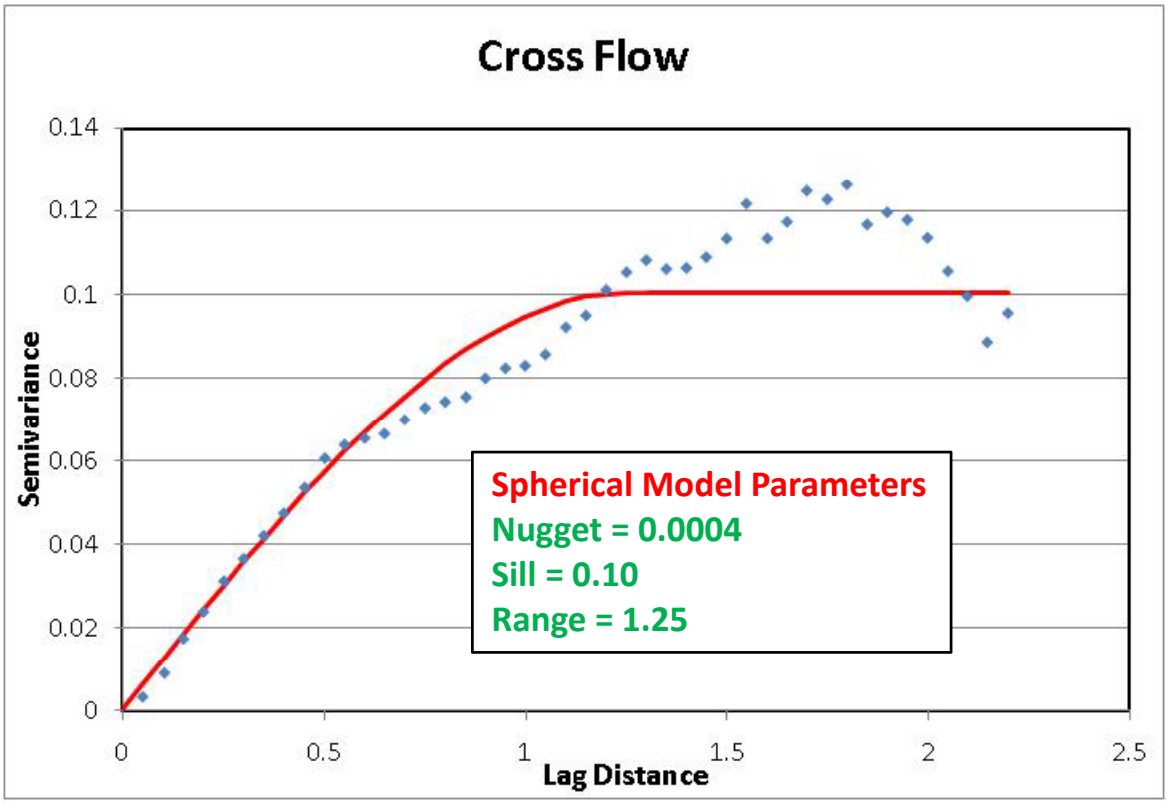
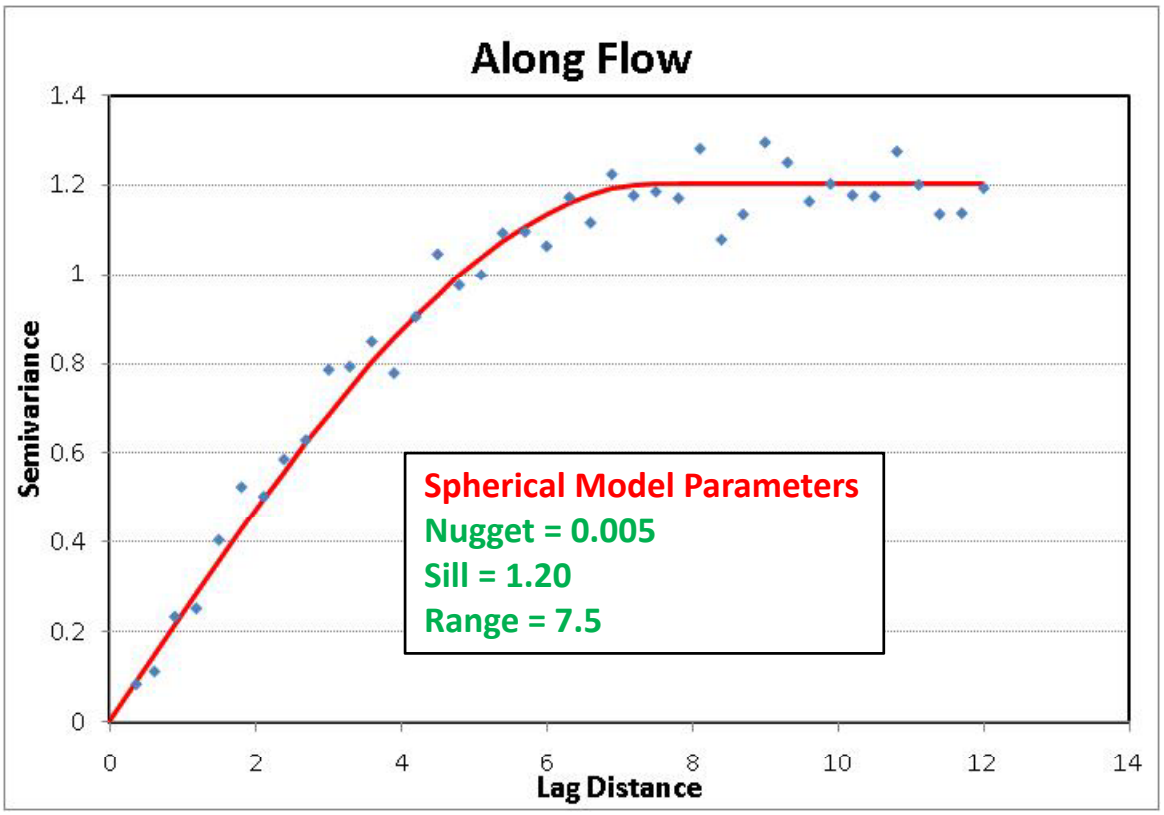


Figure 4k: Directional Semivariograms and Fitted Models for Normal Scores Transformed Residuals of Detrended 2001 Bathymetry



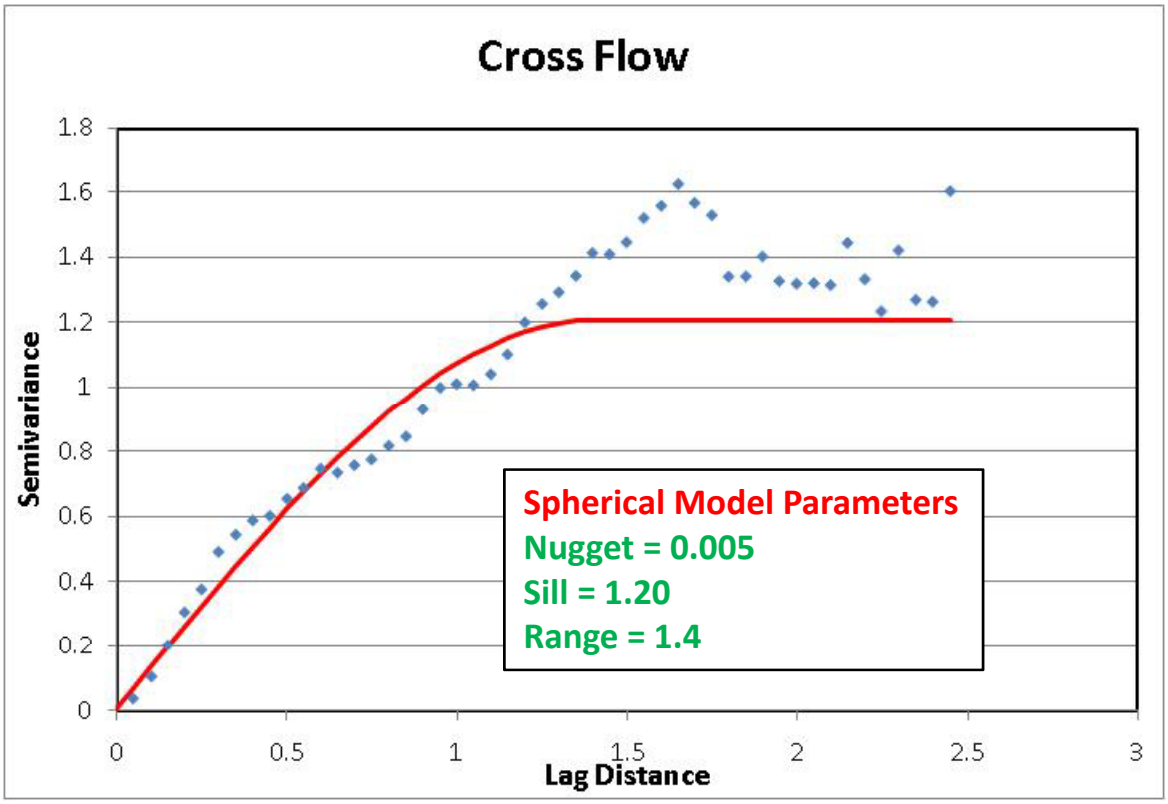
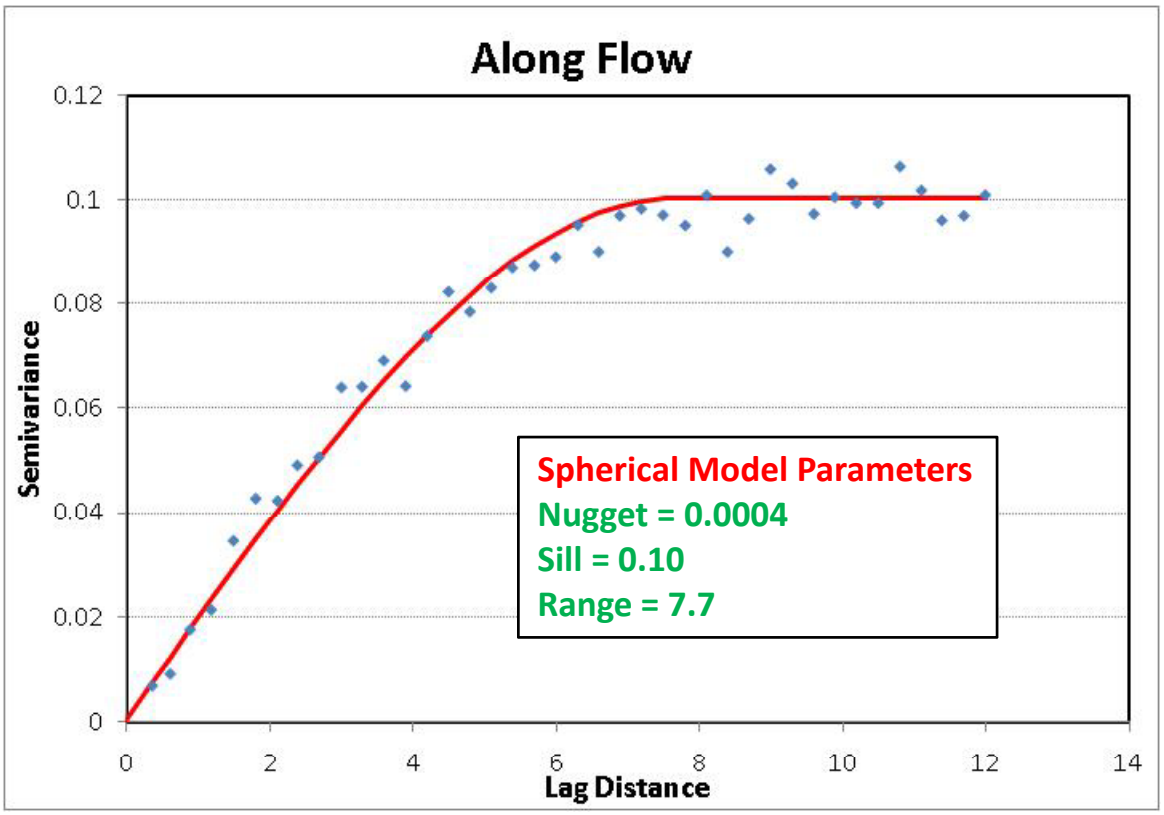


Figure 4l: Directional Semivariograms and Fitted Models for Uniform Scores Transformed Residuals of Detrended 2001 Bathymetry



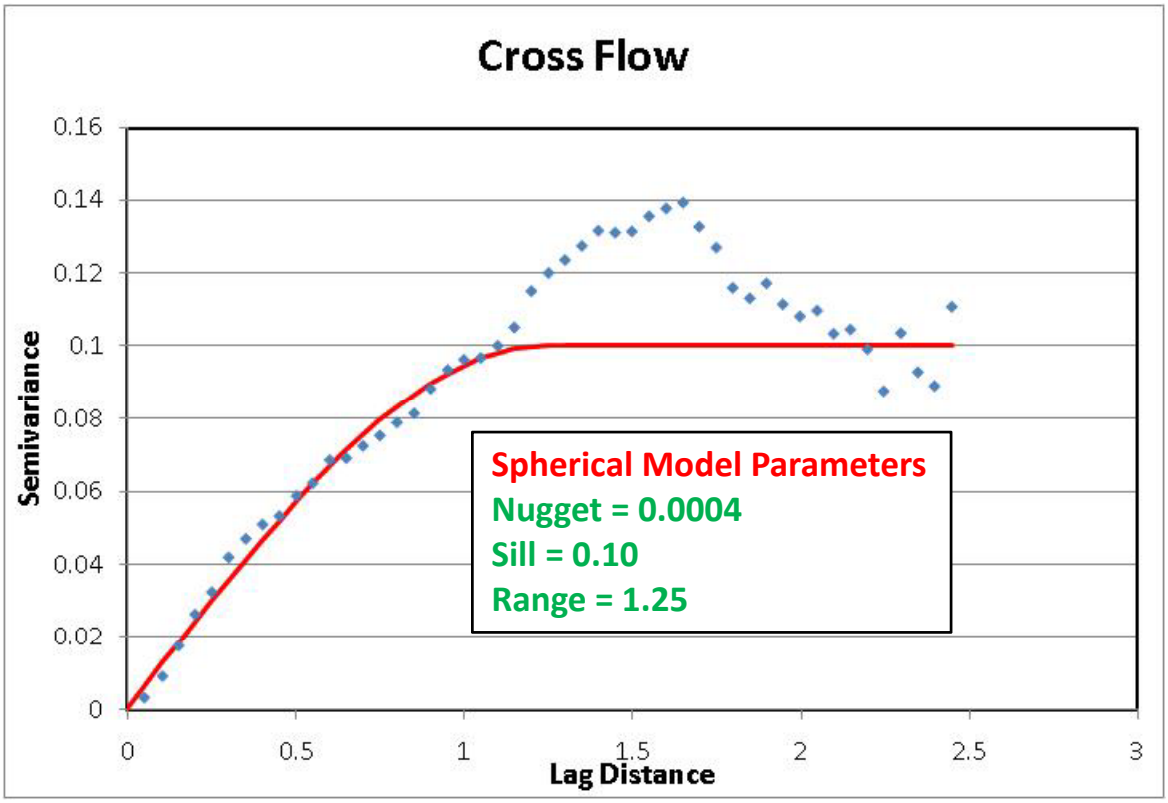
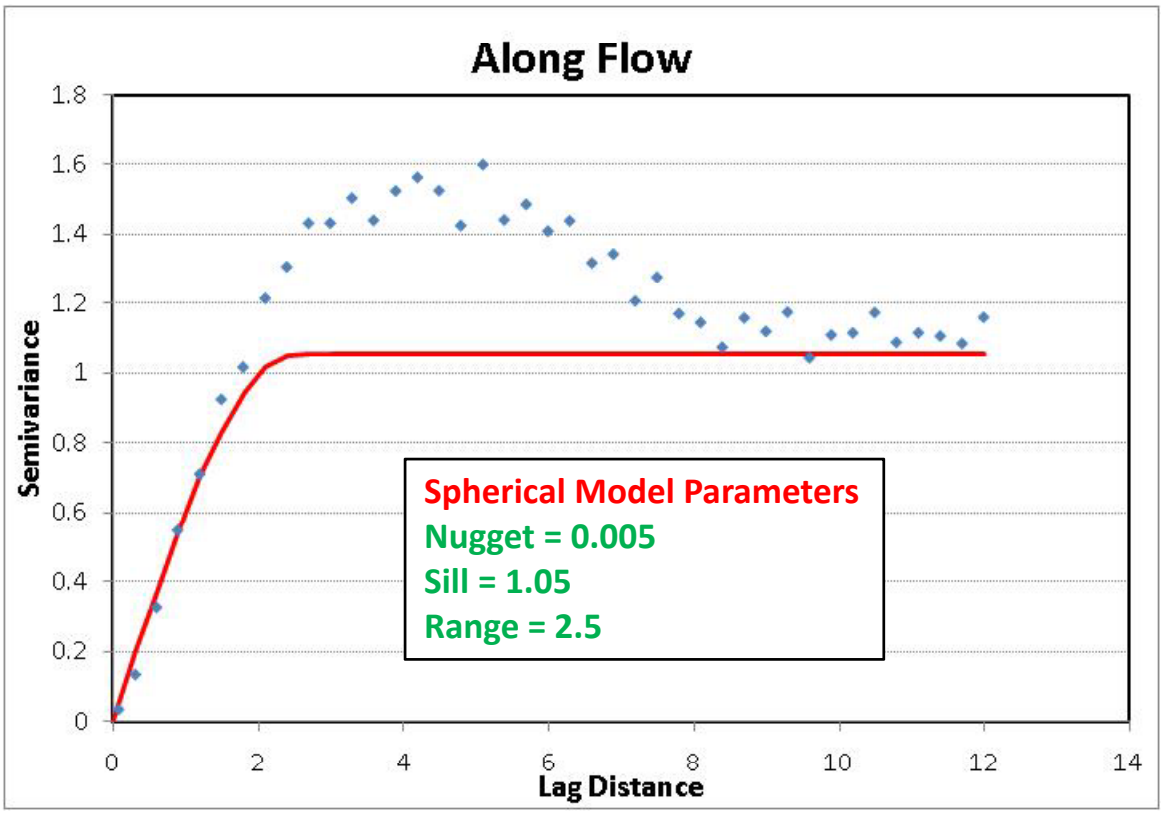


Figure 4m: Directional Semivariograms and Fitted Models for Normal Scores Transformed Residuals of Detrended 2002 Bathymetry



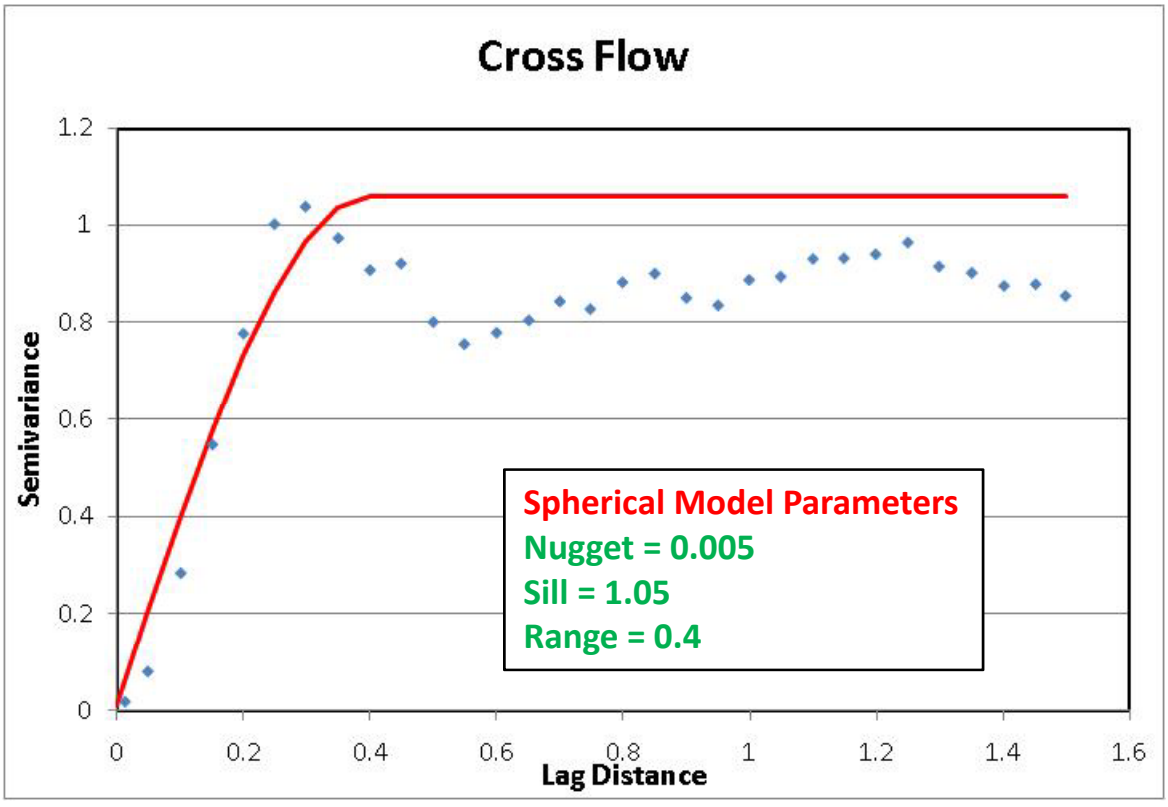
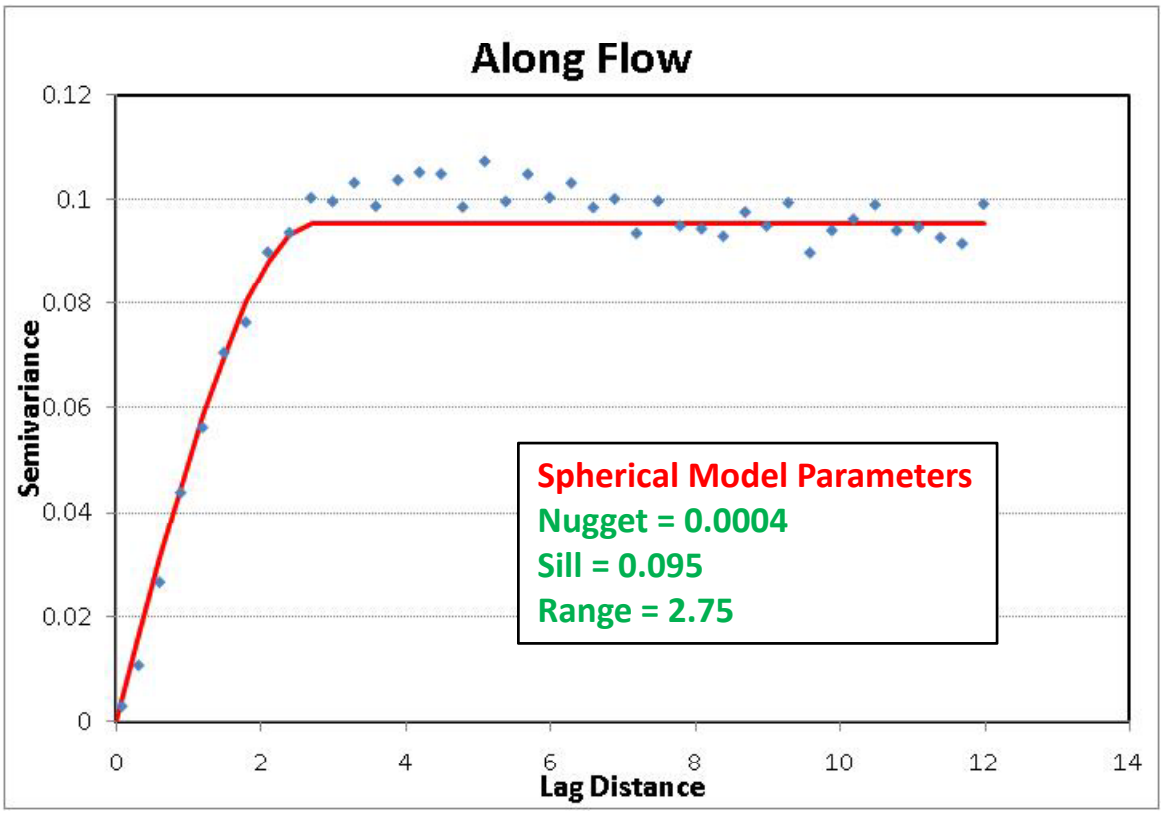


Figure 4n: Directional Semivariograms and Fitted Models for Uniform Scores Transformed Residuals of Detrended 2002 Bathymetry



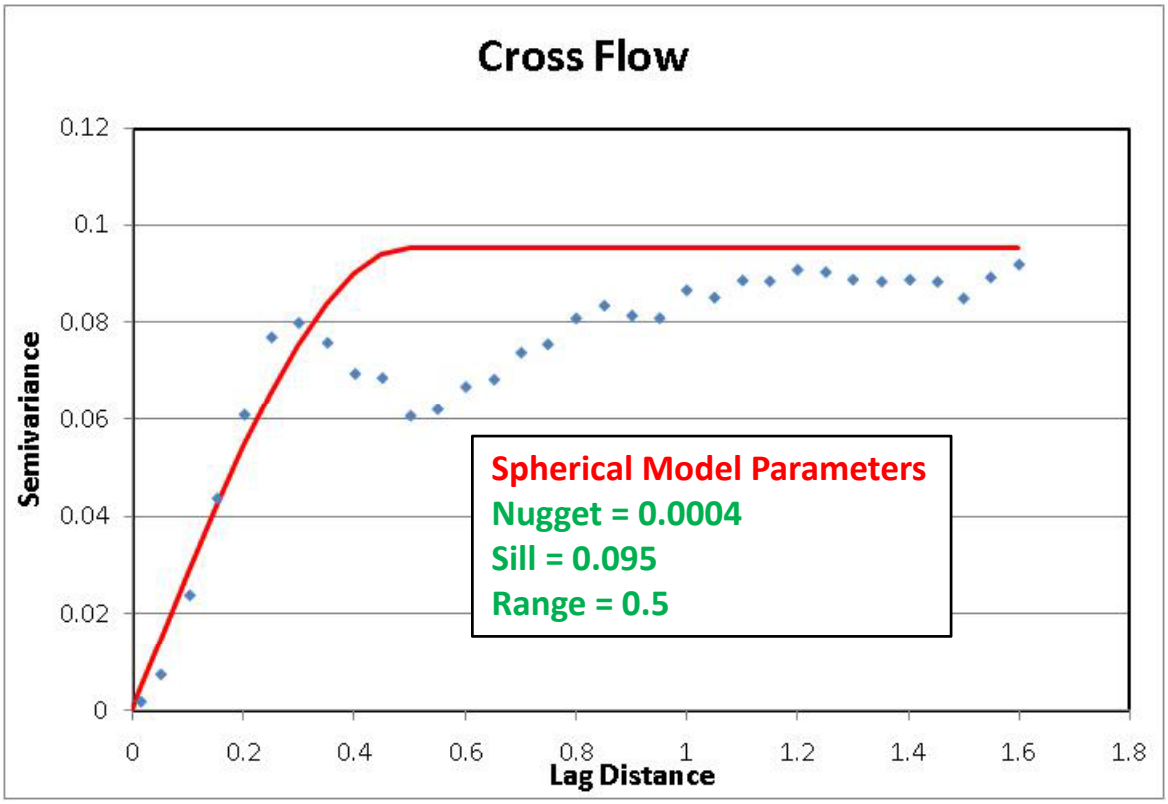
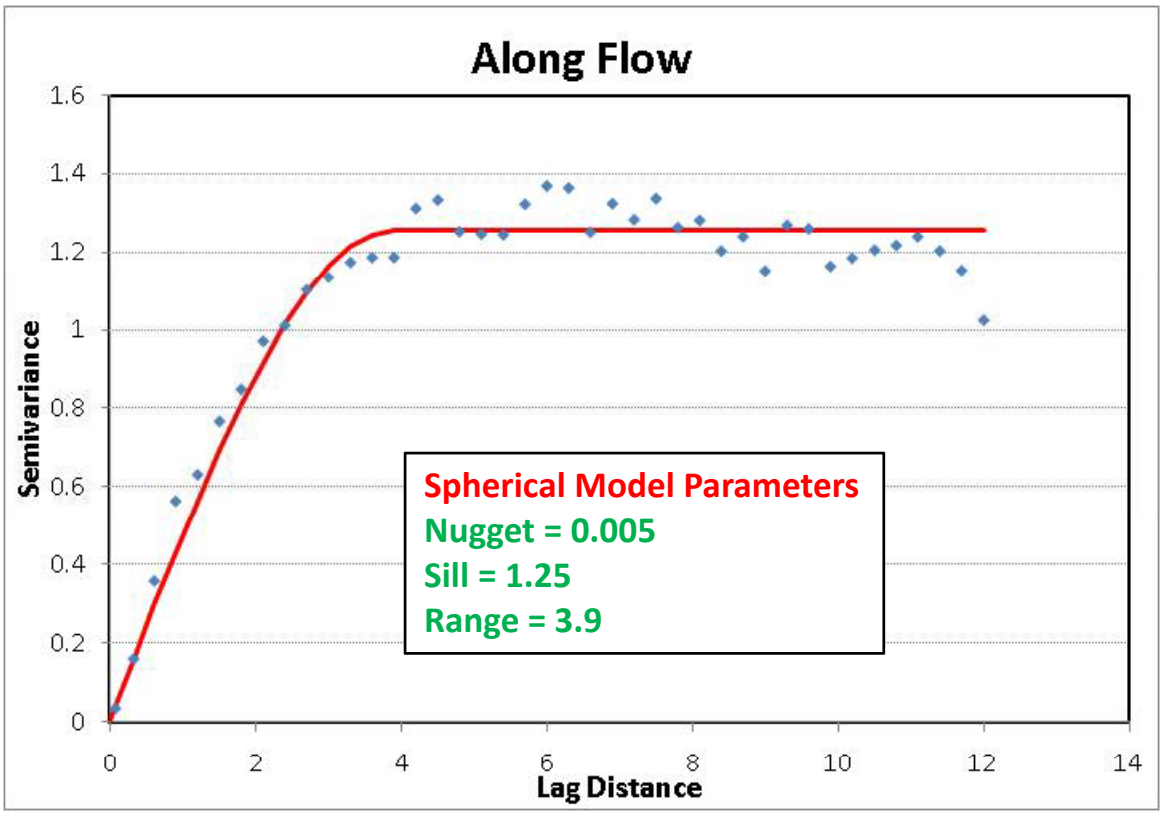


Figure 4o: Directional Semivariograms and Fitted Models for Normal Scores Transformed Residuals of Detrended 2004 Bathymetry



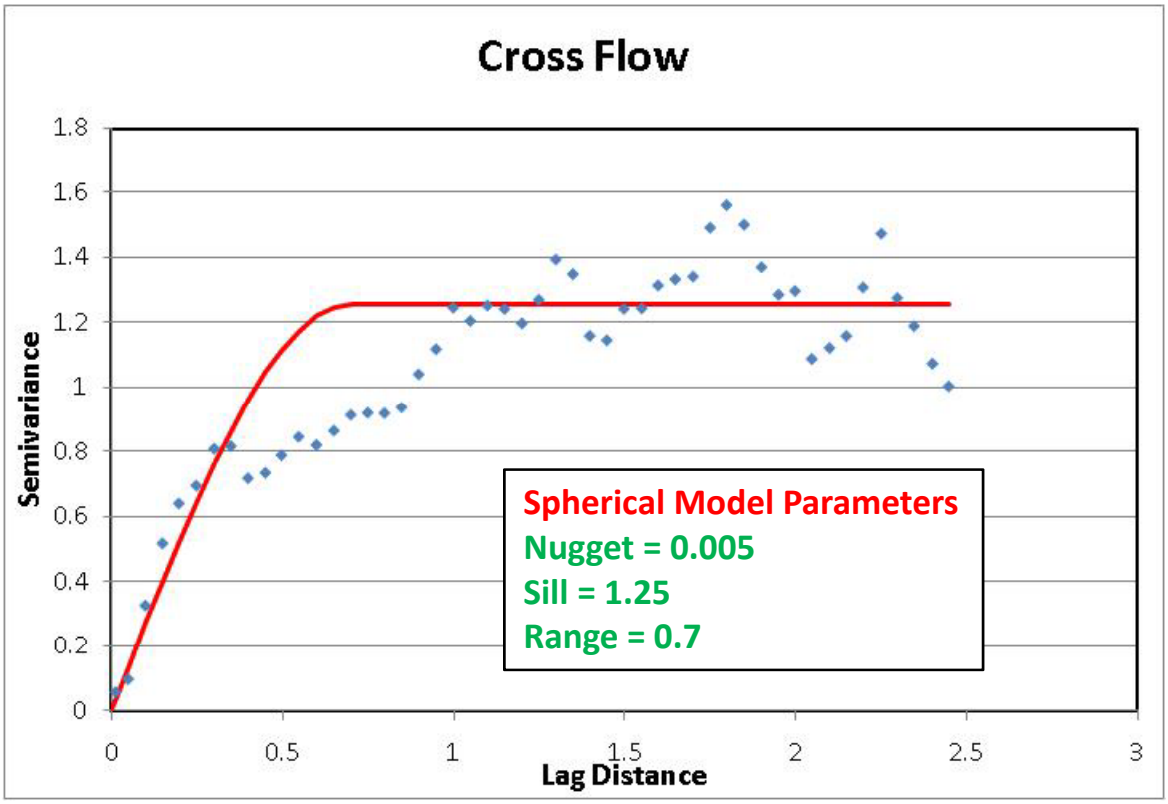
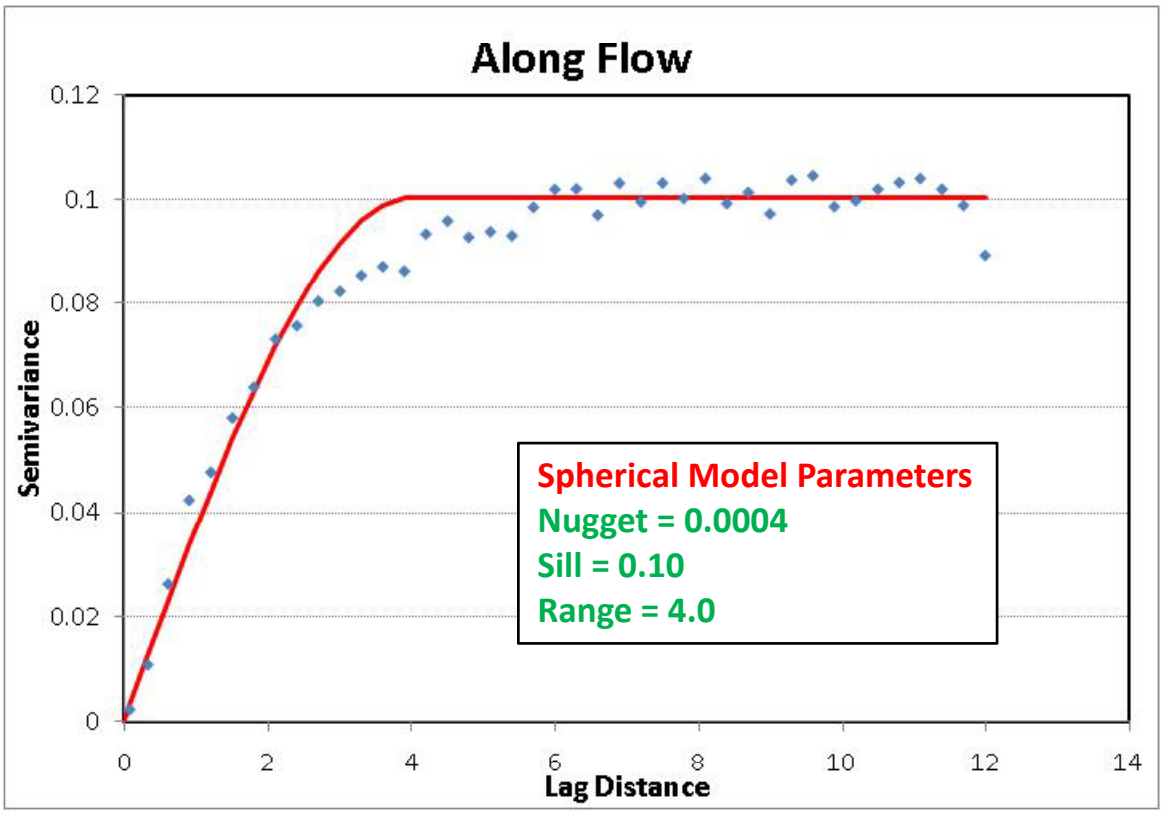


Figure 4p: Directional Semivariograms and Fitted Models for Uniform Scores Transformed Residuals of Detrended 2004 Bathymetry



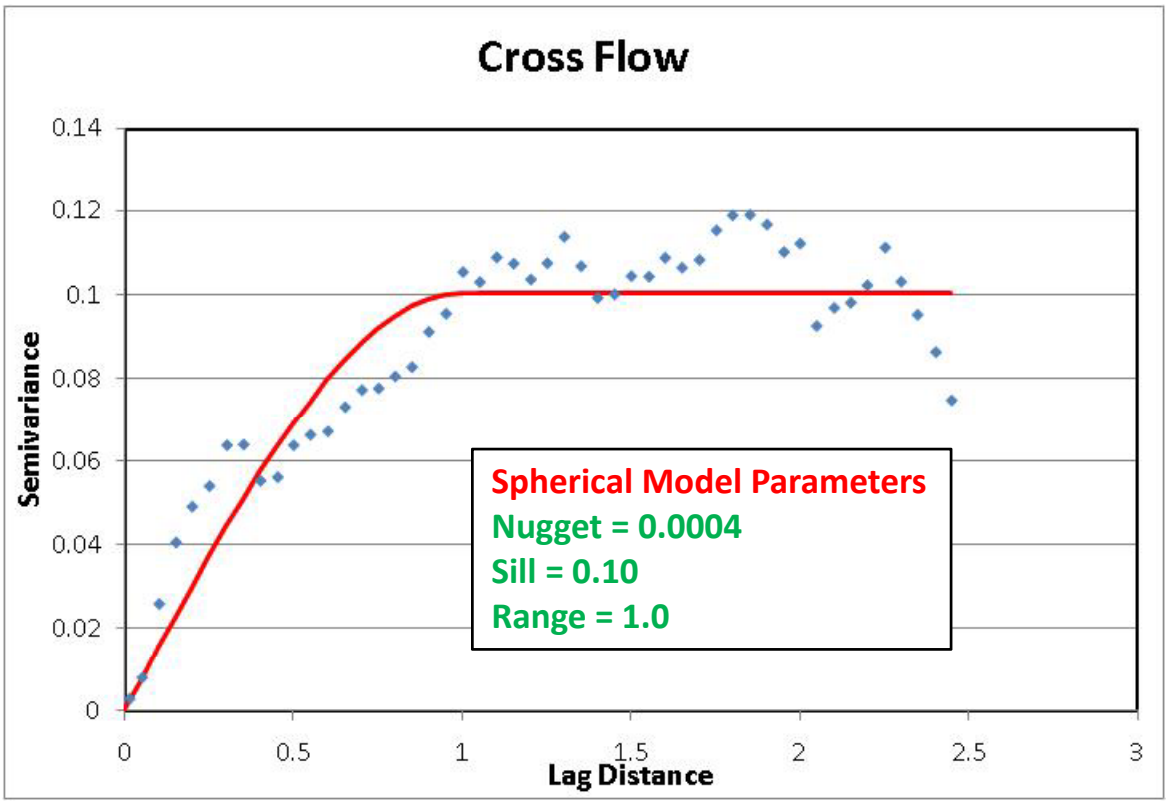


Figure 5: Three realizations of the 1996 bathymetric elevations.

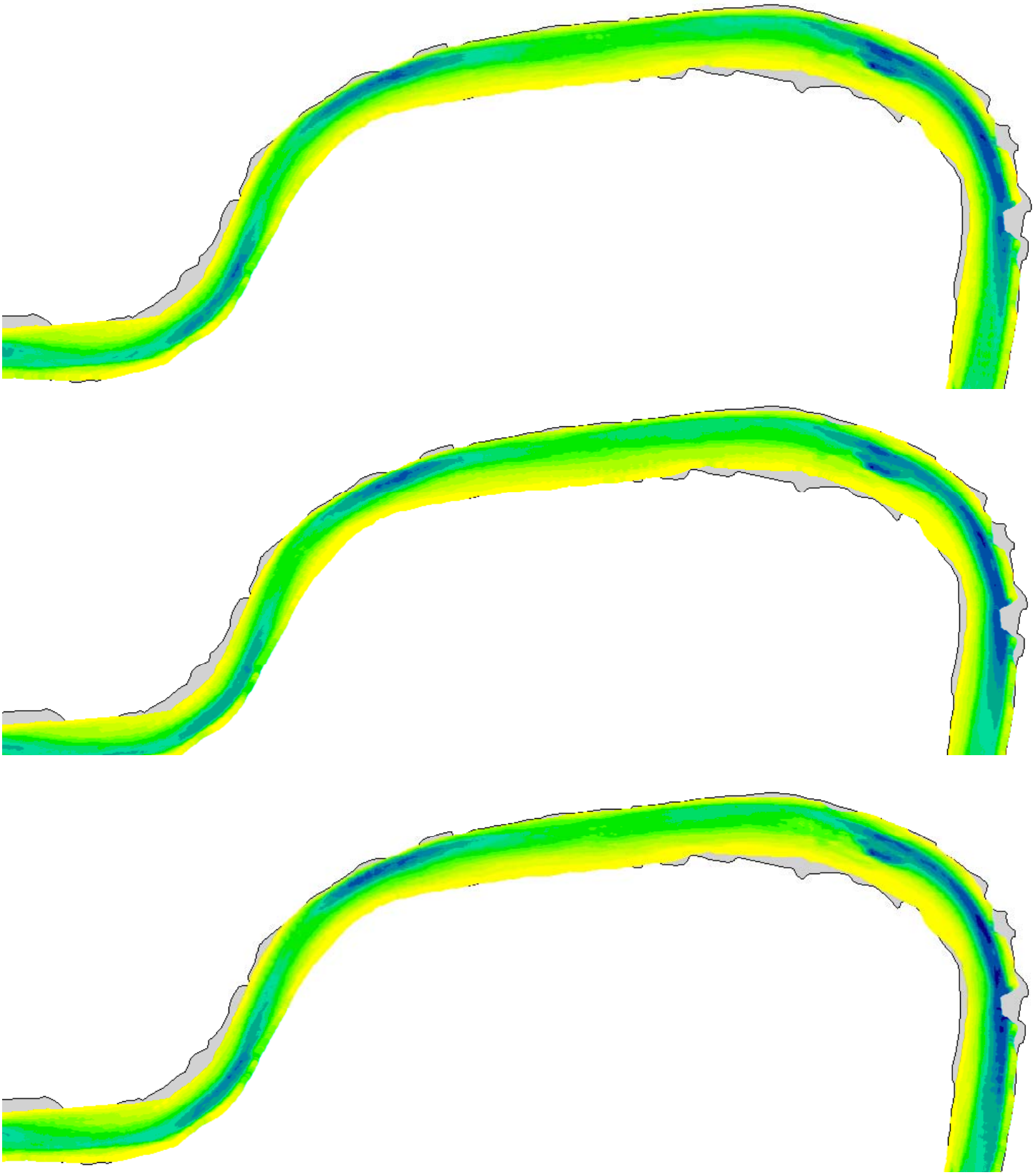
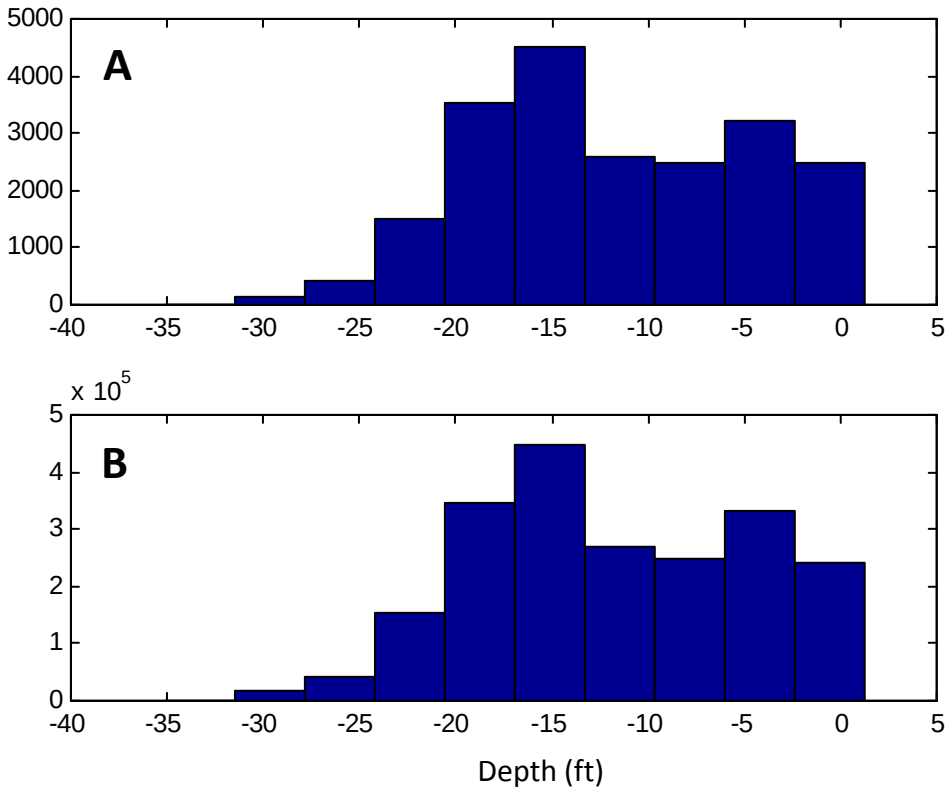


Figure 6: Histogram of sample elevations (Panel A) and simulated elevations (Panel B) for the lower Passaic River in 1995.



Note:

1) Simulated elevations represent all locations at which values were simulated, as opposed to the subset of locations at which inter year comparisons were conducted.
2) This comparison illustrates that the simulation algorithm reproduces the data histogram.
3) Because inter-year comparisons of simulated data were restricted to a smaller lateral extent than the sample data, direct comparison of histograms would be biased toward deeper soundings in the simulated soundings.

Figure 7: Semivariograms for 20 realizations compared with theoretical model semivariograms for cross-flow (green) and long-flow (black) directions.